



Australian Government

Patent Office
Canberra

I, TESSA ROBINSON, EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. 2002952696 for a patent by GENOMICS RESEARCH PARTNERS PTY LTD as filed on 14 November 2002.

WITNESS my hand this
Thirtieth day of June 2009

T. Robinson

TESSA ROBINSON
EXAMINATION SUPPORT AND SALES



Regulation 3.2

Patents Act 1990

for the invention entitled:

"Status Determination"

The invention is described in the following statement:

C:\P\DOCT\Proclaw\FA\Doc\log\G\control\g\control_no revision_final.doc 14/11/02

- 2 -

and often require transport of the animal to a diagnostic centre.

5 A final report of the results of a blood test to an end user, eg. a trainer, often requires involvement of multiple parties each providing separate input to the report. For example, a veterinarian may collect a blood sample, the sample is transported or sent to a laboratory for analysis, personnel in the laboratory perform an analysis using machinery on the blood sample, automated results from the analysis, with or without a veterinary pathologist interpretation, are returned to the veterinarian who then interprets the results and provides a separate report to the trainer. The process is laborious, time consuming, subject to error and interpretation bias and may or may not contain information relevant to the end user.

Bioinformatics may be used with genetic based diagnosis of an animal's health.

15 Currently, it is known to use genetic information in determining information regarding an individual. This can be achieved in a number of ways depending on the information that is desired.

20 Thus, for example, WO 01/25473 describes a method of characterising a biological condition or agent using calibrated gene expression profiles. In this case, when a subject is suspected of having a condition, a test is performed to obtain a specific profile, which is then analysed. In particular, the collected profile is compared to a predetermined profile to determine if the condition has been correctly identified. However, this suffers from drawbacks in that a preliminary diagnosis is required to allow the correct test to be performed.

25 US 6,287,254 describes a system that allows users to perform DNA genetic profiling to determine the susceptibility of a subject to a condition. In particular, in this example, the subject is profiled to determine the presence of predetermined genes, which in turn indicate the susceptibility of a subject to a respective condition. Again, this requires specific tests for specific conditions, and only allows the susceptibility of a subject to be determined.

C:\WDOT\ProDev\VPAL\Dev\op\G\acomb\g\acomb_001\rev\rev_001.doc-14/11/02

- 4 -

The method may further include determining one or more conditions displayed by the subject in accordance with the determined group.

5

The parameters can be representative of the level or abundance of an agent in the subject or in a biological sample obtained from the subject. The agent may be selected from one or more of:

- a) A nucleic acid molecule;
- 10 b) A proteinaceous molecule;
- c) A carbohydrate;
- d) A lipid;
- e) A drug;
- f) A chemical;
- 15 g) A gas;
- h) A cell;
- i) A pathogenic organism; and,
- j) A non pathogenic organism.

20 By "*obtained from*" is meant that a sample such as, for example, a nucleic acid extract or polypeptide extract is isolated from, or derived from, a particular source. For example, the extract may be isolated directly from a tissue or a biological fluid isolated directly from the subject.

25 Other parameters can be measured however, such as the near IR of the subject's blood, general measurements, such as temperature, or other biological indicators.

The method usually includes:

- a) Receiving confirmation of the determined status; and,
- 30 b) Updating the predetermined data in accordance with the confirmed status and the subject data.

C:\W2007\VP\G\VP\A\Davies\Collison\CAVE_Pat.&Trad\Doc-14/11/02

- 6 -

In a fourth broad form the present invention provides a method of allowing a user to determine the status of a subject, the method including:

- a) Receiving subject data from the user via a communications network, the subject data including one or more parameter values, at least one of the parameter being indicative of the current biological status of the subject;
- b) Causing the base station to:
 - i) Compare the subject data to predetermined data, the predetermined data including for each of a number of individuals:
 - (1) One or more parameter values for the respective individual; and,
 - (2) An indication of the status of each individual; and,
 - ii) Determine the status of the subject in accordance with the results of the comparison; and,
- c) Transferring an indication of the status of the subject to the user via the communications network.

The method generally includes:

- a) Having the user determine the subject data using a remote end station; and,
- b) Transferring the subject data from the end station to the base station via the communications network.

The base station can include first and second processing systems, in which case the method can include:

- a) Transferring the subject data to the first processing system;
- b) Transferring the subject data to the second processing system; and,
- c) Causing the second processing system to perform the comparison.

The method may also include:

- a) Transferring the results of the comparison to the first processing system; and,
- b) Causing the first processing system to determine the status of the subject.

In this case, the method preferably includes at least one of:

C:\WDQTV\author\VTADesktop\GCM\cm\gcmctrl_1m revision_1\final.doc 14/11/02

- 7 -

- a) Transferring the subject data between the communications network and the first processing system through a first firewall; and,
- b) Transferring the subject data between the first and the second processing systems through a second firewall.

5

The second processing system may be coupled to a database adapted to store the predetermined data, the method including:

- a) Querying the database to obtain at least selected predetermined data from the database; and,
- 10 b) Compare the selected predetermined data to the subject data.

The second processing system can be coupled to a subject database, the method including storing the subject data in the subject database.

- 15 It is also possible to implement any one of the features of the first broad form of the invention. Thus, for example, the status may include details of any conditions of the individuals, in which case the method can include determining any conditions displayed by the user. The method may also include determining the ability of the subject to perform in a sporting and/or racing event in accordance with any determined conditions.

20

Similarly, the method of performing the comparison typically includes causing the second processing system to:

- a) Obtain a set of templates, the set of templates representing differences between groups of individuals; and,
- 25 b) Use the templates to classify the subject data into a respective one of the groups.

The method can include having the user determine the subject data using a secure array, the secure array having a number of features each located at respective position on the array, and a respective serial number. In this case, the method typically includes causing
30 the base station to:

- a) Determine the serial number from the subject data;

C:\WDQNTV\ProSci\VP\A\Doc\Map\Control\map\map_001\map_001.doc 3/4/1/02

- 8 -

- b) Determine a layout indicating the position of each feature on the array;
 - c) Determining the parameter values in accordance with the determined layout, and the subject data.
- 5 The method may also include:
- a) Receiving confirmation of the determined status from the user; and,
 - b) Updating the predetermined data in accordance with the confirmed status and the subject data.
- 10 In this case, the features can include at least one of:
- a) An oligonucleotide;
 - b) A peptide;
 - c) An antibody;
 - d) A carbohydrate;
 - 15 e) A lipid;
 - f) A cell; and,
 - g) An organism.

The method can also include causing the base station to:

- 20 a) Determine payment information, the payment information representing the provision of payment by the user; and,
- b) Perform the comparison in response to the determination of the payment information.
- 25 In a fifth broad form the present invention provides a base station for determining the status of a subject, the base station including:
- a) A store method for storing predetermined data, the predetermined data including for each of a number of individuals:
 - i) One or more parameter values, at least one of the parameters being indicative of the current biological status of the individual;
 - 30 ii) An indication of the status of the individual; and,

- 30 The processing system can be adapted to:

C:\P\INT\Pro\1411\DAVIES\Collison\CAVE_Pat.&Trad.doc-14/11/02

- 10 -

- a) Obtain a set of templates, the set of templates representing differences between groups of individuals; and,
 - b) Use the templates to classify the subject data into a respective one of the groups.
- 5 The method may further include determining one or more conditions displayed by the subject in accordance with the determined group.

10 The subject data may be determined using a secure array, the secure array having a number of features each located at respective position on the array, and a respective serial number, the processing system being adapted to:

- a) Determine the serial number from the subject data;
 - b) Determine a layout indicating the position of each feature on the array;
 - c) Determining the parameter values in accordance with the determined layout, and
- 15 the subject data.

The processing system can be adapted to:

- a) Receive confirmation of the determined ability; and,
 - b) Update the predetermined data in accordance with the determined ability and the
- 20 subject data.

The base station of the fifth broad form of the invention may therefore be adapted to perform the method of the fourth broad form of the invention.

25 In a sixth broad form the present invention provides a computer program product for determining the status of a subject, the computer program product including computer executable code which when executed on a suitable processing system causes the processing system to perform the method of the fourth broad form of the invention.

30 In a seventh broad form the present invention provides an end station adapted to determine the status of a subject, the end station including a processor adapted to:

- 12 -

- ii) An indication of the status of each individual;
- c) Determining the status of the subject in accordance with the results of the comparison; and,
- d) Providing an indication of the ability in accordance with the results of the comparison.

5

The method of determining the status of the subject may be the method of the first or fourth broad forms of the invention.

- 10 The status of each individual typically indicates any conditions displayed by the user, in which case the method typically includes:
- a) Determining any conditions displayed by the user in accordance with the results of the comparison; and,
 - b) Determining the ability in accordance with the determined conditions.

15

In a tenth broad form the present invention provides apparatus for determining the ability of a subject to perform in a sporting and/or racing event, the apparatus including a processing system adapted to:

- 20
- a) Obtain subject data, the subject data including one or more parameter values, at least one of the parameter being indicative of the current biological status of the subject;
 - b) Compare the subject data to predetermined data, the predetermined data including for each of a number of individuals:
 - i) One or more parameter values for the respective individual; and,
 - ii) An indication of the status of each individual;
 - c) Determine the status of the subject in accordance with the results of comparison; and,
 - d) Provide an indication of the ability in accordance with the results of comparison.

25

30

- 13 -

The processing system is generally adapted to perform the method of the ninth broad form of the invention.

In an eleventh broad form the present invention provides a computer program product for
5 determining the ability of a subject to perform in a sporting and/or racing event, the computer program product including computer executable code which when executed on a suitable processing system causes the processing system to perform the method of the ninth broad form of the invention.

10 In a twelfth broad form the present invention provides a method of providing secure arrays for use, each array including a number of predetermined features, the method including:

- a) Determining a number of respective feature layouts, each layout representing the positioning of each feature on a respective array;
- b) Determining a number of serial numbers, each serial number corresponding to a
15 respective layout;
- c) Generating a number of arrays, each array being generated in accordance with a respective layout, and including the corresponding serial number thereon, the serial number being used in processing the array.

20 The method can be performed to provide the arrays on behalf of an entity, the method including providing an indication of the layouts and corresponding serial numbers to the entity, to thereby allow the entity to process the arrays.

The method of determining the layouts typically includes:

- 25 a) Determining a preferred layout; and,
- b) Moving the position of one or more of the features from the position in the preferred layout to alternative position.

The method can include:

- 30 a) Determining the type of each feature; and,
- b) Exchanging the position of one or more features having different feature types.

C:\P\DOC\ProDoc\7A\Doc\7A\Gens\7A\prodoc\7A\revisions_001.doc 14/11/02

- 14 -

Brief Description of the Drawings

An example of the present invention will now be described with reference to the accompanying drawings, in which: -

- 5 Figure 1 is a schematic diagram of an example of a processing system for implementing the invention;
Figure 2 is a flow chart outlining the process implemented by the system of Figure 1;
Figure 3 is a schematic diagram of an example of a distributed architecture for implementing the invention;
- 10 Figure 4 is a schematic diagram of an example of one of the end stations of Figure 3;
Figures 5A and 5B are a flow chart of the process implemented by the system of Figure 3;
Figure 6 is a flow chart of the process for generating templates;
Figures 7A and 7B are a flow chart of the process of comparing the subject data to the templates;
- 15 Figure 8 is a schematic diagram of a second example of a distributed architecture for implementing the invention;
Figure 9 is a flow chart of the process for generating secure arrays; and,
Figure 10 is a flow chart of the process for generating subject data using the secure arrays.
Figure 11 is a flow chart of the process of data mining;
- 20 Figure 12 is a flow diagram illustrating dataflow steps in a specific example as part of a computer system capable of delivery of remote diagnostic services;
Figure 13 is a flow diagram showing steps for diagnosing a condition of an animal in accordance with a specific example;
Figure 14 is a diagram illustrating an environment for working the specific example shown
25 in Figure 13;
Figure 15 is a flow diagram illustrating steps for preparing an array in accordance with a specific example of the invention;
Figure 16 is a flow diagram showing steps for determining a nucleic acid expression level in a biological sample; and
- 30 Figure 17 is a flow diagram illustrating steps for building a database in accordance with a specific example.

C:\WPCVT\PreG2\177\ALD\alup\Orcas\alup\gradmk_001\preG2_001.doc 14/11/02

- 15 -

Detailed Description of the Preferred Embodiments

An example of the present invention will now be described with reference to Figure 1, which shows a processing system suitable for implementing the present invention.

5

In particular, Figure 1 shows a processing system 10 including a processor 20, a memory 21, an optional input/output (I/O) device 22 and an interface 23 coupled together via a bus 24. In use, the interface 23 is adapted to couple the processing system 10 to one or more databases shown generally at 11.

10

In use, the processing system 10 is adapted to receive subject data, which is data representative of the current biological status of a subject. The subject data is typically in the form of raw data and therefore requires interpretation to allow the status of the subject to be determined. This is achieved by having the processing system 10 compare the subject data to predetermined data stored in the database 11. The predetermined data includes data representative of the biological status of a number of individuals, together with an indication of the actual status of the individuals when the predetermined data was collected.

15

Accordingly, by comparison of the subject data with the predetermined data, this allows the subject data to be interpreted and the current biological status of the subject to be determined.

20

The manner in which this may be achieved will now be described in outline with respect to Figure 2.

25

In particular, at step 100 the user determines subject data in the form of parameter values representing the current biological status of the subject. In particular, the parameter values represent specific measurements of selected parameters that represent the current biological status of the subject. It will be appreciated that a number of different forms of parameters may be used, as will be described in more detail below.

30

C:\PCT\TUP\Doc\17\AID\English\Documents\2000\2000_141102

- 16 -

At step 110 the user provides the parameter values to the processing system 10, which then operates to compare the subject data to the predetermined data at step 120. In particular, the predetermined data includes parameter values for a number of individuals having a range of different biological states.

By comparing the subject and predetermined data this allows the processing system 10 to determine the status of the subject in accordance with the results of the comparison at step 130. Thus, the processing system 10 will attempt to identify individuals having similar parameter values to the subject. The status of the subject will then be determined to be similar to that of the identified individuals.

Once the status has been determined the processing system 10 provides an indication of the status to the user at step 140.

This procedure can therefore be used to identify a wide range of conditions that may be displayed by the subject. In particular, the system can be adapted to determine the presence of one or more of a number of conditions in the subject.

In order to achieve this, each of the number of conditions must have been previously identified in the individuals. Accordingly, it is necessary to have predetermined data for a large number of individuals, with at least some of the individuals having a range of the conditions. Furthermore, it is also necessary to utilise a sufficiently large number of parameters to allow the respective conditions to be distinguished on a statistical basis.

The parameters used and typical numbers will be described in more detail below. However, it will be appreciated that the number of parameters required will generally increase depending on the number of conditions being identified.

Accordingly, it is typical for the predetermined data to include values for a large number of parameters and individuals.

C:\WDOFT\WDOFT\VPAN\Debug\Control\Algorithm_00\version_00\Lev014\1102

- 17 -

It will therefore be appreciated that determination of the predetermined data is typically a time consuming and expensive procedure. It is not therefore feasible for each user wanting to implement the method to collect their own predetermined data. Accordingly, the invention is typically implemented using a distributed processing system an example of which is shown in Figure 3.

As shown, in Figure 3 the apparatus is formed from a base station 1 coupled to a number of end stations 3 via a communications network 2, and/or via a number of LANs (Local Area Networks) 4. The base station 1 is generally formed from one or more of the processing systems 10 coupled to a data store, such as the database 11, as shown.

In use, the processing system 10 operates substantially as described above to process data received via the communications networks 2, 4. The processing system 10 can then supply an indication of the determined subject status back to the respective end station 3 via the communications network 2, 4, as will be understood by a person skilled in the art.

In use, this allows the base station to be administered by an operator, that provides services allowing users of the end stations 3 to determine the status of a subject. This in turn overcomes the need for each user to obtain their own predetermined data. Furthermore, by having the base station 1 perform the comparison of the subject and predetermined data, and determine the status, this allows the operator of the base station 1 to restrict access to the predetermined data, thereby preventing the data being accessed and used by third parties. This, in turn allows the operator to charge a fee for the provision of an indication of the status of the subject, as will be described in more detail below.

In any event, it will therefore be appreciated that the system may be implemented using a number of different architectures. However, in this example the communications network 2 is preferably the Internet 2, with the LANs 4 representing private LANs, such as LANs within a company or the like.

C:\PDOC\7\ProGm\77A\NucDry\Control\procmk_01\revision_001.doc 14/11/02

- 19 -

more detail with reference to Figures 5A and 5B.

In particular, as set out in Figure 5A the process begins at step 200 with the user determining the parameter values for the subject. The parameter values are then encoded
5 as subject data by the end station 3 at step 210. This is typically achieved in accordance with a predetermined algorithm such that the subject data has a predetermined format that can be interpreted by the base station 1.

At step 220, the user accesses the base station 1 using the end station 3.
10

At this stage, the user of the end station 3 will typically be required to either register with the base station 1 or supply a predetermined user name and password. In particular, this is performed to allow the base station 1 to determine the identity of the user and therefore confirm that the user has authorisation to utilise the services provided by the base station 1
15 and/or to ensure that payment can be obtained for the provision of the services.

It will be appreciated that the user name and password will typically be provided when the user registers with the base station 1 on a first occasion. At this point the user has to make provisions for payments, such as the provision of account details, thereby allowing the
20 operator of the base station 1 to charge the user for the services provided.

The user name and password will then be generated and subsequently verified in the normal way. Alternatively, identification of the user can be achieved in accordance with cookies stored at the end station 3, or an identifier associated with the end station 3, which
25 may for example be the MAC (Media Access Control) address of the end station interface 33, or the like.

In any event, it will be appreciated that access to the services provided by the base station 1 is generally limited to authorised users.
30

In any event, when the user accesses the base station 1, this is typically achieved by

C:\WP007\Work\VPAD\Doc\of\Documents\processes_as_written.doc 14/11/02

- 20 -

accessing respective web pages generated by the base station 1. This allows the user to select the respective services required, which in this example is an indication of the status of a subject.

- 5 Once the user has been authorised, the user will be transferred to a secure environment to allow the subject data to be transferred to the base station 1 for processing. This is typically achieved by implementing an SSL (Single Socket Layer) connection between the base station 1 and the end station 3. This provides additional security and in particular, to ensure that the subject data transferred between the base station 1 and end station 3 is
- 10 retained confidential.

Confidentiality of the subject data and the determined status are important as the results are often used in determining the ability of the subject to compete in sporting and/or racing events, this information can be extremely valuable, especially to the gambling industry. It

15 is therefore necessary to ensure the information is retained confidential at all times.

After accessing the base station 1 at step 220, the subject data is transferred to the base station 1 at step 230. At this point, the base station 1 will typically operate to review the subject data to ensure that it is genuine subject data, and that for example, the data does not

20 disguise an attempt to hack into the base station 1 to obtain access to the predetermined data. This is typically achieved by having the base station 1 implement a firewall between the processing system 10 and the Internet 2 or LANs 4 to ensure that unwanted data is not received.

- 25 In any event, at step 240 the processing system 10 operates to determine the nature of the subject data.

Thus, it will be appreciated that the exact subject data provided and, in particular, the parameters for which values are provided may vary depending on the respective

30 implementation. This will be described in further detail below. However, it will be appreciated that the subject data may be collected using arrays, in which case a number of

- 21 -

different arrays may be provided. Thus, in this case, the base station 1 will operate to determine the type of array being used, to allow the subject data to be interpreted.

5 At step 250 the processing system 10 selects at least some of the predetermined data in accordance with the nature of the subject data. Thus, for example, the processing system 10 will operate to select parameter values from the predetermined data for parameters corresponding to those contained in the subject data.

10 At step 260 the processing system 10 compares the parameter values of the subject data to the parameter values of the selected predetermined data. In particular, the processing system 10 operates to compare the parameter values to those obtained from a number of different individuals that between them have a range of different conditions. This allows the processing system 10 to determine one or more conditions displayed by the subject at step 270.

15

At step 280 the processing system 10 optionally determines the ability of the subject to compete in a sporting and/or racing event in accordance with the determined conditions. The processing system 10 then transfers an indication of at least the conditions to the end station 3 at step 290.

Thus, it will be appreciated that the system may be implemented in a variety of ways. Typically however the subject data is formed from phenotypic information representative of the current biological status of the subject. In one embodiment, the phenotypic information results from the expression of the genotype of the subject and is therefore typically in the form of information such as expression data, or the like. The expression data may relate to the level or abundance of an RNA molecule or a polypeptide. The RNA molecule includes, but is not restricted to, RNA transcripts such as a primary gene transcript or pre-messenger RNA (pre-mRNA), which may contain one or more introns, as well as a messenger RNA (mRNA) in which any introns of the pre-mRNA have been excised and the exons spliced together, heterogeneous nuclear RNA (hnRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), small cytoplasmic RNA

C:\WORK\Patent\WPA\Hepato\Case\Ingrained..._revision_06.docx 34/11/02

- 22 -

(scRNA), ribosomal RNA (rRNA), translational control RNA (tcRNA), transfer RNA (tRNA), eRNA, messenger-RNA-interfering complementary RNA (micRNA) or interference RNA (iRNA) and mitochondrial RNA (mtRNA). Suitable polypeptides that are contemplated by the present invention include enzymes, receptors, immunoglobulins, hormones, cytokines, chemokines, neuropeptides, adhesins, glycoproteins and the like. Alternatively, the expression data may relate to the level or abundance of a carbohydrate including monosaccharides, oligosaccharides and polysaccharides.

When the phenotypic information relates to expression data, these are typically obtained by any suitable qualitative or quantitative technique. However, where it is necessary to determine the level or abundance of a multiplicity of different expression products, it is preferable to use multiplexed analysis techniques including arrays of distinctly detectable beads as is well known in the art.

The subject data may optionally contain genotypic information including genetic information carried in the chromosomes and extrachromosomally. Such data may be obtained from genetic mapping, genetic screening, pedigree, family history and heritable physical and psychological characteristics.

In this case it will be appreciated that the expression data collected may be relevant to a respective condition that is already diagnosed in the subject. However, advantageously the present invention can be utilised to detect previously undiagnosed conditions. In particular, this can be achieved by collecting sufficient parameter values and then comparing these to the predetermined data which is being collected for individuals having a range of conditions. This then allows conditions to be identified before symptoms are necessarily visible.

This can therefore be used in situations such as diagnosing conditions in animals. This is particularly advantageous as the animals are unable to provide information regarding any conditions from which they may be suffering. Thus, in the case of race horses for example, race horses can suffer from a number of conditions, such as overtraining,

C:\WD\NTU\user\1VPA\Desktop\ControlLogbook_# revisions_6-14-11.m2

- 23 -

respiratory illness, or the like, which can be difficult to detect. In contrast to a human athlete, who can usually communicate any symptoms to a trained medical practitioner, horses are unable to communicate to vets and therefore can only be examined passively. Accordingly, the present invention allows a vet or other medical practitioner to perform an analysis of the subject and in particular their current biological condition and determine whether the subject is suffering from any conditions.

10 However, it will be appreciated that this is also useful for diagnosing conditions in humans, where the human may not be aware of the condition. This is particularly the case with high performance athletes where a minor condition may not be noticeable to the athlete directly, but may have an impact on the athlete's performance.

15 The system is also useful for diagnosing conditions in situations where the athlete is trying to keep the condition secret, for example, in the case of drug testing to detect banned substances used by the athlete.

In order to be able to identify a significant number of conditions successfully, it is necessary to have a statistically significant quantity of predetermined data. In particular, it is necessary to have predetermined data obtained from one or more individuals suffering from a respective condition to allow the condition to be identified, and the sample size will therefore have to be sufficiently large to ensure this occurs. For example, if the chance of an individual from a general population having a specific condition is 1 in 100, it will be necessary to sample at least 100 individuals to ensure at least one individual having the condition is sampled. In fact, it would in this case be typical to sample at least 1000 individuals, to ensure that one or more individuals having the condition are identified.

Furthermore, the more data available from individuals suffering from the condition the better as this allows distinctions to be drawn between individuals suffering from different types of conditions.

The number of parameters required will depend on the number of conditions to be

C:\WINNT\Program Files\DAVIES COLLISON CAVE\Pat. & Trad\14-11-02

- 24 -

distinguished. In particular, it will depend on factors such as:

- The presence and detection of unknown conditions;
- The range of conditions to be identified within the population;
- The levels of incidence of each condition in the population;
- 5 • The ability to distinguish between the conditions.

It will be appreciated that as individuals, including performance animals such as race horses, can suffer from a wide variety of conditions, then it is preferable for a large number of parameters such as 3,000 to 5,000 to be used. However, this number can be
10 significantly lower if only a minor number of conditions are to be identified. Thus for example, the number of parameters used may be anywhere from 10 up to 10,000, or more. Suitably, the number of parameters employed are at least about 20, preferably at least about 50, more preferably at least about 100, even more preferably at least about 150, even more preferably at least about 200, even more preferably at least about 300, even more
15 preferably at least about 500, even more preferably at least about 1000, even more preferably at least about 1500, even more preferably at least about 2000, even more preferably at least about 4000, even more preferably at least about 6000, even more preferably at least about 8000, and still even more preferably at least about 10000.

20 In addition to this, the effect of a condition on an individual may also vary in accordance with additional phenotypic information relating to a particular characteristic or set of characteristics of the subject, as determined by interaction of the subject's genotype with the environment in which it exists. In this embodiment, such 'characteristic data' may be selected from age, sex, height, length, weight, ethnicity, race, breed of animal, feeding
25 patterns, exercise patterns, medication supplied, nutritional or growth supplements supplied, nutritional analysis, hair colour, skin colour, eye colour, body composition, fat composition, water retention, obesity, transcriptomic profile, proteomic profile, blood type, tissue type, endocrine function, immunological function including cellular and humoral immune function, tolerance, allergy, transplant rejection, cancer, hyperplasia,
30 gastrointestinal function, neurological function, kidney function, heart function, brain function, pancreatic function, bone function, joint function, sexual or reproductive

C:\WD\NT\Pro\Doc\VP\AD\ad\wp\General\general_001.rvt100-1411403

- 25 -

function, metabolic load, toxicological profile, substance abuse including drug dependency, inborn errors of metabolism, infectious disease including viral infection, bacterial infection, mycobacterium infection, parasitic infection, prion function, prosthesis, tissue reconstruction, surgery, pain and the like. The phenotypic information may also include demographic information, which can be important for monitoring the spread of a condition globally, as well as to allow analysis to take account of conditions that are limited to predetermined areas. Thus, it is generally preferable to additionally collect characteristic data together with the expression data for the individuals.

10 It will be appreciated that in view of the amount of predetermined data involved, it is not generally feasible to compare subject data determined for any one subject to all the predetermined data. Accordingly, some pre-processing of the predetermined data is generally performed. This process is generally known as data mining and will be described in more detail below and will now be outlined with reference to Figure 6.

In particular, at step 300 the operators of the base station 1 operate to collect the predetermined data, including genotypic and phenotypic information, from a number of individuals.

20 In this example, it will be assumed that the parameter values that form the predetermined data correspond to expression data and in particular, concentration quantities, abundances, or ratios of respective expression products obtained from an array or the like. The phenotypic information will typically be provided from a study of the respective individual, and is preferably provided in a standard format to allow the information to be

25 correctly interpreted by the base station 1.

At step 310 the processing system 10 operates to generate parameter vectors for each individual. Each parameter vector is formed from a vector containing each of the parameter values at a respective location within the parameter vector. The processing system 10 then operates to consider the position of the parameter vectors in an N-dimensional space, where N corresponds to the number of parameters at step 320.

C:\W0071\BrdGen\YPA\Del\q\Gen\cond\gen\cond_14\11\02

- 26 -

At step 330, the processing system 10 operates to determine groups of the individuals in the predetermined data. In particular, the groups may be identified by considering individuals having similar physical conditions.

5

Thus, the processing system 10 will operate at step 340 to produce diagnostic templates that are used to characterise the groups of individuals identified in the predetermined data, and to classify individuals into these groups. It will be appreciated that there is a multiplicity of ways of defining such diagnostic templates for example regularised discriminant analysis, Support Vector Machines, recursive partitioning, artificial neural networks, or the like, as will be described in more detail below with respect to data mining.

Having identified the groups, and the templates, the processing system 10 will operate at step 350 to characterise the ability of the diagnostic templates to predict group membership, by applying the templates to the individuals comprising the predetermined data. This characterisation and validation may be achieved, for example, by k fold crossvalidation, and the construction of permutation distributions.

Once the templates have been generated, these are then stored in the database at step 360, together with any other summary statistics necessary to provide statistically efficient prediction using the template.

It will be appreciated that the definition of groups may be done for respective individuals irrespective of the phenotypic traits. Thus, for example, if all individuals suffering from a condition tend to have similar parameter values, then all the individuals having the condition will be contained in the same group irrespective of each individual's phenotypic traits.

However, if different phenotypic types have distinct parameter values for the same condition, then a respective template will be defined for each phenotypic group. Thus for example, a template may be defined for male horses having a respiratory condition, with a

C:\PQNT\A\of\an\7A\1\Descript\Descript\gencat22_00000000_001.doc 14/11/02

- 27 -

separate template being defined for female horses having the same respiratory condition.

In addition to this, at least one group will correspond to healthy individuals not having any conditions. Thus a template will be generated corresponding to a healthy individual. It will be appreciated that this can be used in determining if an individual has an unidentified condition, as will be described in more detail below. This can also be used to identify sub-clinical diseases or conditions that are not previously apparent through existing diagnostic techniques.

Once the templates have been generated, it is then possible to operate to compare the subject data to the predetermined data. The manner in which the comparison is performed will now be described with reference to the flow chart shown in Figures 7A and 7B.

In particular, at step 400, the user determines parameter values, and phenotypic information relating to the subject. At step 410, the end station 3 is used to generate subject data in accordance with the determined parameter values and phenotypic information. At step 420 the user transfers the subject data to the processing system 10 as described above.

At step 430, the processing system 10 extracts the parameter values and the phenotypic data from the subject data, and then uses the parameter values to generate a parameter vector at step 440.

The processing system obtains one or more of the templates from the database at step 450. At this point, it will be appreciated that the templates may be selected in accordance with the phenotypic information, such that the subject parameter vector is only compared to templates having suitable phenotypic traits. Thus, for example, it will be appreciated that if the subject is a male horse, then it may be pointless comparing the subject parameter vector to a template representing a group of female horses having a respiratory disease.

However, if a template corresponds to a group of individuals having a range of phenotypic

C:\w\071\Project\W7A\Doc\ep4\Gen\ep4\gen\pat_1_001.doc - 14/11/02

- 28 -

traits, then this template will be used to predict group membership using the subject parameter vector, at step 460. It will be appreciated that there is a multiplicity of ways of predicting group membership from the subject parameter vector, just as there is a multiplicity of ways of constructing group templates, as will be appreciated by persons skilled in the art.

At step 470 the processing system 10 operates to determine the uncertainties in group prediction using the subject parameter vector and templates in the N dimensional vector space. These uncertainties are expressed as probabilities that the test subject has a condition previously characterised by membership of one of the groups in the predetermined data.

It is apparent to those skilled in the art that there is a multiplicity of ways of constructing these uncertainties, each appropriate for a different method of template construction and group prediction. For example uncertainties may be based on some measure of distance between the subject parameter vector and a group template, or by a Bayes rule applied to a set of discriminant functions.

It will be appreciated therefore that the templates may be based on specific values such that they represent a single point in the N dimension vector space. Alternatively however the templates may correspond to ranges such that each template defines a range of parameter values for which the subject would have the respective condition. Thus, this effectively defines decision boundaries in the N dimensional space, such that if the subject parameter vector falls within the decision boundary, this indicates that the subject has the respective condition.

If the parameter vector is approximately equidistant to two or more templates, this may indicate that there is a chance that the individual either has a previously undetermined condition, or alternatively is suffering for example from a combination of the two conditions. It will be appreciated that templates may be generated for common combinations of conditions, as well as single conditions.

C:\P\N\T\U\T\O\G\I\V\A\I\Descript\G\en\re\te\gradual_26\rev\descrip_Gen\Des-1471162

- 29 -

Finally, it will be appreciated that the presence of the template for healthy individuals allows a healthy subject to be determined. If the subject parameter vector is significantly separated from this template, this will indicate that the subject is generally unhealthy, and this allows previously unidentified conditions to be determined, for example, if the subject parameter vector is not near any of the other templates.

It will also be appreciated that the magnitude of the parameter values will allow the severity of conditions to be determined. Thus, for example, the greater a difference in magnitude between the parameter values for a healthy subject compared to a subject suffering from a condition will generally indicate a greater severity of the respective condition.

Similarly, it will be appreciated that groups may be defined for different severity of condition. Thus, for example, a first group may be defined for the initial stages of a condition that is treatable, whilst a second group is defined for the same condition when it has progressed beyond the initial stages and is no longer treatable.

Finally, a direct comparison of the subject parameter values can be made with the predetermined data for other individuals suffering from the same condition, can also be used to allow the severity of the conditions to be determined.

In any event, at step 480, the processing system 10 interprets the separation of the parameter vector from the templates and uses this to determine any conditions displayed by the subject. An indication of this is then transferred to the end station 3 at step 490.

A number of alternatives can be implemented in the present invention.

Multiple Firewall

In particular, in the above described example it will be appreciated that users of the end stations 3 are unable to access any of the data stored in the database 11. This is performed

CONFIDENTIAL/PAID/Do not distribute outside services facilities-1471303

- 31.-

Once the comparison has been performed, the processing system 14 can determine those conditions suffered by the subject and then transfer an indication of this back to the processing system 12 through the firewall 15. The processing system 12 can then transfer an indication of this indication to the end station 3.

It will be appreciated that in this example even assuming the user is able to infiltrate the first firewall 13, the user will only be able to access previously submitted requests and the results determined therefrom. The presence of the dual firewall system therefore makes it virtually impossible for the user to infiltrate the processing system 14 and obtain access to the data stored in the database 11.

In the remainder of the description, it will be appreciated that the processing systems 10; 12, 14 are effectively interchangeable.

Parameter Ranges

A further alternative to the present invention is for the comparison to be performed on the basis of parameter ranges defined for different conditions.

20 Thus for example, each condition may have associated therewith a sequence of parameter value ranges determined based on ranges of parameter values for individuals diagnosed with the respective condition. The parameter value ranges can then indicate for a respective condition the parameter values that can be expected, allowing the determined parameter values to be compared to the respective range for each condition to determine if

25 the parameter values provided fall within a respective range.

Thus, a respective parameter range can be determined for each condition, with the parameter values determined for a subject being compared to each range, to determine those ranges within which the subject data falls.

30 An indication of the likelihood of the subject having a respective condition can then be

CSWDCVTU Profile V.P. Aldesha, G. Ganeshaiah, et al. as reviewers. Final Doc. 14/1/2021

- 32 -

determined statistically based on the number of individuals having the respective condition.

Multi-Level Analysis

5 In the example described above, a number of conditions have been defined for the respective type of individual. However, it will be appreciated that sometimes it is desirable to perform tests to focus on specific conditions. Thus for example, in the case in which a horse has an existing condition, it is sometimes desirable to monitor the development of the condition for the respective subject.

10

In this case, as the condition has been determined, it will not usually be necessary to consider all of the parameter values each time the analysis is performed. In particular, as a large number of parameters are provided to allow the different conditions to be distinguished, a large number of parameters will typically not be representative of the progress of a specific condition.

15

Thus, it is usually possible to identify a number of key parameters that are relevant to respective conditions. Thus for example, conditions relating to respective respiratory illnesses may be uniquely identified using a smaller number such as 50 parameters. In this instance, if the user is only interested in examining for the progress of this respective condition, the user can simply supply an indication of the values for the respective 50 parameters.

20

25 In this example, the processing system 10 would operate to compare the determined parameter values against parameter values of horses suffering from the condition and horses not suffering from the condition. In this situation, the manner in which the collection of the parameter values is performed may vary.

25

30 In the examples described above it has been mentioned that the parameter values may include for example expression data collected using an array, for example. If the arrays are to collect values corresponding to 5000 parameters it is typical for an array to be provided

30

C:\P\NT\Prod\KPA\Verap\10\unsub\pat\ca\ca_141102_141102.doc

- 33 -

with 5000 features thereon with each feature corresponding to a respective parameter. Alternatively, 10,000 features may be provided with two features corresponding to each parameter. In any event, a person skilled in the art will appreciate that a number of variations on this are possible.

5

However, if only 50 parameters are to be measured, it is then possible to provide an array having 5,000 features with 100 features being used to determine the value for each parameter. This allows the parameter values to be determined far more accurately allowing a more accurate representation of the condition to be determined. In particular, more accurate comparison of the subject data with the predetermined data can be performed.

Thus, a typical sequence of events may be for a user to submit a general test having a large number of parameters similar to that described above which allows respective conditions to be first identified. Once a condition has been identified, the user can then purchase specifically designed array plates adapted to monitor the specific condition. Measurements of the parameter values relevant to the condition can then be made far more accurately allowing the progress of the condition to be monitored in detail. This can allow users to be provided with information concerning whether conditions are improving or not.

20

Longitudinal Analysis

In the methods described above, the subject data for a respective subject is compared to predetermined data for a number of different individuals. However, in addition to, or alternatively to this, longitudinal analysis can also be performed. In this instance, the subject data is compared to subject data previously collected for the same subject. Thus, this allows the progression of a condition within a subject to be monitored.

25

Again, it will be appreciated that if this is performed with a limited number of parameters as described in the multi-level analysis described above, then this allows an accurate assessment of the progression of a condition to be made.

30

C:\WP51\TheGenVP\A\Doctryp\Control\figs\cont_11 rev1.doc, final.doc 14/11/02

- 34 -

By storing the results determined for a respective subject in the database 11, for a predetermined time period, this can allow the progression of the disease over a time period to be monitored and displayed to the user. Thus, the most recently obtained subject data is compared to earlier subject data for the same subject (and optionally predetermined ata), to
5 determine disease progression.

Thus, for example, levels of respective parameter values can be used to indicate the severity of the disease. This can be achieved by comparing the subject data to predetermined data in the manner described above, or alternatively using other techniques.
10 As the parameter values vary over time, this can be used to provide an indication of whether the condition is improving or worsening. This in turn can be used to monitor the effectiveness of any treatment given to the subject.

Thus for example, if it is determined that a horse has been overtrained then the obvious
15 solution to this problem is to reduce training for a predetermined time period, or resting the horse. However, trainers will generally not want to reduce the training too much as the horse will become unfit. Similarly, worse problems can arise if the trainer resumes training too early.

20 Thus, in this case, the trainer can submit subject data on a periodic basis such as every week allowing the fitness of the horse to be determined on a weekly basis. An indication of this can then be transferred back to the user allowing the trainer to determine when training of the horse should resume, or how hard training should be.

25 This therefore allows the severity of the condition within the subject to be monitored.

Secure Arrays

Arrays, or genechip technology determines whether genes are turned "on" or "off" under certain conditions. A single gene is a stretch of DNA found in the nucleus of a cell that
30 encodes, through mRNA, the information required to produce a particular protein. Genechips measure the level of mRNA in cells and can be directly correlated to the

In the example of collect subject data for horses, a horse array and a blood sample are needed. The array has DNA dotted onto its surface (DNA of the genes in horse blood cells). The DNA on the array consists of one strand of the double-stranded DNA molecule – the other strand is provided by the blood sample and is labelled with a dye.

Two strands of similar DNA will only bind to each another (hybridise) if they match in...
sequence. An array reader can determine the amount of mRNA in a sample (gene turned
10 "on" or "off") by determining the amount of dyclabelled DNA that hybridises to an array.

The reader produces a value compared to a reference for every single gene on the array. The 5,000 to 10,000 values can then be compared to Genetraks' database. Genes turned "on" or "off", individually or in patterns, can then be identified and correlated to the specific conditions of a racehorse.

Various conditions in racehorses will alter the metabolism of the white blood cells, which can then be detected using the genechip technology. For example, the gene for manganese superoxide dismutase (MnSOD) may be turned "off" in respiratory inflammation. 20 Similarly, IFNg, IL-4 may also be turned "off", and the genes for Grola, IL-8, TNF and MIF may be turned "on". This pattern of "gene expression" can be correlated to a specific condition, such as respiratory inflammation caused by a virus. Patterns of gene expression change as a horse succumbs to or recovers from a viral infection. As the technology and database develops, predictions on the stage of infection or influence of treatments can be 25 made.

As described above, it is preferably to ensure that the predetermined data is retained as confidential.

30 However, if arrays are used in the collection of data, it will be appreciated that it would be possible to purchase a quantity of arrays and perform data mining of data obtained from

C:\WP\PROF\DAVIES\Alderbrook\new\log\patent_log\log_141102

- 36 -

used arrays to determine new predetermined data. Thus, there is a danger that competing companies will use the arrays provided on behalf of the operator for their own purposes.

5 In order to be able to do this, it will be necessary for the competing entities to be able to interpret the data provided by the arrays. However, this can be overcome by utilising secure arrays. In particular, secure arrays utilise a randomisation of the layout of the array to avoid the problems of reverse engineering or the like.

10 The manner in which this may be achieved will now be described with reference to Figures 9 and 10.

In particular, at step 600 the operators of the base station 1 will determine a number of features to be included on the array, and provide an indication of these features to the array supplier at step 610.

15 At step 620, the array supplier will operate to generate a preferred array layout using a processing system. This is performed in accordance with normal operating procedures. In particular, the array suppliers will generally utilise applications software to determine a preferred array layout which optimises the array build process. The layout is generally
20 organised so that creation of the array is simplified.

At step 630, the array supplier will operate to generate a number of randomised array layouts. The randomised array layouts have one or more of the features positioned in
25 alternative location when compared to the preferred array layout. In particular, the array supplier will generally operate to move or swap the locations of one or more of the features on the array. In order to swap features, it must be ensured that the features are of different types.

At step 640 the array supplier will also operate to generate a corresponding number
30 serial numbers. In particular, a respective serial number is provided for each randomised array layout that is generated.

C:\W2017\ProSeis\VA\DeSeis\Oce\lab\gsd\ndt_10\rev\ndt_6\ndt_6_14/11/02

- 37 -

At step 650 the array supplier will operate to generate arrays in accordance with the randomised array layouts and the serial numbers. In particular, each generated array will have features positioned thereon in accordance with a respective one of the randomised layouts, together with an indication of the corresponding serial number.

It is typical for the array supplier to produce the arrays in batches with up to 1,000 arrays in each batch, with each batch being created in accordance with a difference randomised layout.

10

At step 660 the randomised arrays are transferred to the users for subsequent use in generating the subject data, whilst at step 570 the serial numbers, together with corresponding layouts are transferred to the base station 1.

15 The use of randomised arrays will slightly complicate the production process but will vastly increase the security of the arrays. In particular, third parties will be unable to utilise the arrays, as the location of features alter, which will cause the third parties to obtain varying results on different arrays, for the same sample.

20 Operation of the system to use the randomised arrays will now be described with reference to Figure 10. In particular, at step 700 the user will obtain a biological sample from the subject and then perform an assay process using the array at step 710.

At step 720 the user uses the end station 3 to encode the values obtained from the array as subject data, together with a serial number indication. The subject data is then transferred to the base station 1, in the manner described above, at step 730.

The processing system 10 operates to determine the serial number from the subject data at step 740. The serial number is then used to access the respective array layout stored in the database 11 at step 750.

C:\W\ONT\Work\DAV\A\DeskTop\Ocarina\Ocarina_20 rev1.doc 6/11/02

- 38 -

The array layout will then be used by the processing system to interpret the subject data, and in particular, to determine the respective feature to which each value corresponds. This allows the processing system 10 to hence determine the parameter values for the respective subject data.

5

Operation of the invention will then be substantially as previously described above.

It will be appreciated that the serial number may also be used to check the user is an authorised user. In particular, if each user is provided with arrays having a respective serial number (a range of serial numbers), then having the array supplier provide an indication of the user and the serial number(s) to the operator, this allows the operator to verify the identity of the user. This provides an audit trail for the arrays.

Feedback

15 A further way in which the present invention may be utilised is to provide feedback on the accuracy of provided results.

In particular, if the base station 1 is used to provide an indication of one or suspected conditions in a subject, the user can be requested to provide an indication whether the diagnosis provided by the base station 1 is correct. This may form a requirement, such that a user will only be provided with services by the base station if they agree to this term.

In any event, the correctness of the assessment by the base station 1 can usually be determined by either treating the subject and determining if the treatment is successful, or by monitoring the development of the condition over a predetermined time period. Once it has been determined that the diagnosis is correct or incorrect, an indication of this can be transferred to the base station 1.

At this point, the respective subject data collected for the respective subject can be saved as predetermined data in the database 11, with the confirmation of the condition being used as the indication of the condition in the predetermined data.

C:\WINDOWS\system32\cmd.exe /c cd C:\Program Files\Google\Google Toolbar\; dir

- 39 -

In order to achieve this, the processing system 10 is typically coupled to a sample database that is used to store the subject data obtained from each subject. Once confirmation of the conditions is received the subject data and the condition indication is transferred to the predetermined data stored in the database 11.

It will be appreciated that this checking of the conditions is not essential to the present invention as typically the data alone will be useful. However, checking of the condition will be useful in determining the accuracy of the templates.

It will be appreciated that as further data is collected over through the feedback technique or through the use of alternative data collection methods the templates or other data can be updated allowing more accurate condition analysis to be performed.

15 Users

It will be appreciated that any individual may use the system. Initially at least however, it is necessary for the user to be able to generate the subject data. In the case in which arrays are used, for example, this requires the user to first collect biological material, such as blood, and then analyse the material using the array. This is generally difficult and requires skilled operators using existing technology. Accordingly, the user may have to be a skilled technician. However, it is envisaged that collection techniques will become simpler, allowing the process to be implemented by any user.

In the case of sporting or racing events, for example, the users could include:

- 25
- Athletes;
 - Trainers;
 - Drug testing committees (such as Olympic Officials);
 - Medical practitioners (such as vets or doctors);
 - Event organisers;
- 30
- Pathology labs (that would typically perform the work on behalf of an individual, such as a horse owner).

C:\R2007\VP\A\Doc\up\Genom\reports\at review_final.doc 14/11/02

- 42 -

stations 1 may be provided at respective geographical locations. This can increase the efficiency of the system by reducing data bandwidth costs and requirements, as well as ensuring that if one base station becomes congested or a fault occurs, other base stations 1 could take over. This also allows load sharing or the like, to ensure access to the system is
5 available at all times.

In this case, it would be necessary to ensure that each database 11 contains the same information and templates such that the use of different ones of the base stations 1 would be transparent to the user.

10

Subject Data

The subject data may be selected from any expression product of the genome or characteristic or set of characteristics of the subject whose levels or abundance may vary within the subject or between two or more different subjects depending on their status. The
15 data include, but are not restricted to, biological, physiological and pathological data of the subject. Examples of biological data include, transcriptomic profiles, proteomic profiles, enzyme function, receptor function, and the like. Physiological data may be selected from age, sex, height, length, weight, ethnicity, race, breed of animal, feeding patterns, exercise patterns, medication supplied, nutritional or growth supplements supplied, hair, skin and
20 eye colour, fat composition, obesity, blood type, tissue type, endocrine function, immunological function, gastrointestinal function, neurological function, kidney function, heart function, brain function, pancreatic function, bone function, joint function, prosthesis, tissue reconstruction, surgery, pain and the like. Examples of pathological data include infectious disease including viral infection, bacterial infection, mycobacterium
25 infection, parasitic infection, prion function, cancer, transplant rejection, inflammatory diseases such arthritis and fibrosis, toxicological profiles, substance abuse including drug dependency and the like.

Data Mining

30 Data Mining

C:\P\WOT\Profile\VF\AD\Debug\Oscilloscope\jcs_noridm3_000102

- 43 -

The system uses a self learning classification system, in which diagnosis is made using a historical database of test results (the predetermined data), which is updated as each test sample (subject data) is recorded. The historical database is typically maintained on a server.

5

Clinical application of the system can be used to diagnose a subject such as an animal with an unknown clinical or performance state. That is, the animal may or may not have some disease, or may or may not be race-ready. A metabolic profile is measured for the animal subject.

10

In a preferred example, the metabolic profile is comprised of expression signatures measured on an oligonucleotide chip. In a preferred example the metabolic profile is compared with a set of pre-computed diagnostic signatures (templates), and together these are used to predict the health status of the subject. In a preferred example, prediction will include probabilistic estimates of uncertainty, and be accompanied by a list of possible differential diagnoses.

15

Diagnostic signatures are computed by data mining a historical database, which contains metabolic profile data on subject animals (predetermined data), and associated clinical information on subject health and performance status. These historical data use the same metabolic profile measurement technique as is used in clinical application. In a preferred example, these metabolic profiles are comprised of expression signatures measured on an oligonucleotide chip.

20

25 Data mining may be performed using a number of techniques including:

- Regularised discriminant analysis for high dimensional data, as described by Kiiveri (1992) Canonical variate analysis of high dimensional spectral data. Technometrics 34 pp. 321-331.
- Diagonal discriminant analysis as described by S. Dudoit, J. Fridlyand, and T. P. Speed (2002). Comparison of discrimination methods for the classification of

30

- 44 -

- Support Vector Machines as described by M P. S. Brown, W Noble Grundy, D Lin, N Cristianini, C Sugnet, T S. Furey, M Ares, Jr., D Haussler (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proceedings of the National Academy of Science*. 97(1):262-267. and Y. Lee, Y. Lin, and G. Wahba (2002) Multicategory Support Vector Machines, Theory, and Application to the Classification of Microarray Data and Satellite Radiance Data. Technical Report 1064. Department of Statistics, University of Wisconsin-Madison.
- Bayesian kernel fitting algorithms.
- Tree based recursive partitioning Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984) *Classification*, augmented by Bagging Breiman, L. (1996) Bagging predictions, *Machine Learning* 26(2) pp. 123-140 and Boosting Breiman, L. (1998) Arcing classifiers. *Annals of Statistics* 26(3) pp. 801-849

20

Diagnostic signatures are combined with test subject metabolic profiles to produce a diagnosis. In one example, (where data mining was based on regularised or diagonal discriminant analysis), prediction is based on a Bayes classification rule, and estimates of uncertainty are based on posterior probabilities of class membership.

25 In another example, (where data mining is based on Support Vector machines),
classification is based on the support vectors, and uncertainties are estimated from distance
of the test profile to the decision boundary. In another example (where data mining is
based on recursive partitioning) classification is based on the estimated decision tree, or
30 averaged over multiple decision trees.

C:\WINNT\Program Files\Adobe\Acrobat\Acrobat.exe

- 45 -

It will usually be the case that even for an animal with an unknown clinical condition or performance status, some clinical or performance conditions are known. For example, it may not be known whether or not the animal has disease A, but it *is* known that the animal has disease B and does not have disease C. When test samples are recorded, the historical database is updated to include the test sample, and any known concomitant clinical or performance information.

It will usually be the case that an animal is tested more than once during a period of investigation. Re-testing may occur at a time when an earlier unknown clinical condition has become known. For the example given above, it may be the case that at a time of re-testing for race-readiness it is known that during the initial test the animal *did* have disease A. Provision is made to allow updates to and modification of the clinical data obtained for each test subject, as diagnosis is confirmed or modified.

In one example, data mining is repeated at regular intervals as the historic database grows. Test records added to the historic database will frequently contain only partial clinical or performance data. For any given clinical or performance factor, data will be filtered to remove subjects for which the particular characteristic is unrecorded. The data mining algorithm will then be used to construct new diagnostic signatures for the given clinical or performance characteristic. The procedure of filtering and mining is repeated for each characteristic of interest. In this way, the sample sizes used to obtain diagnostic signatures are constantly increasing, and predictive performance improves. The system becomes self-learning.

The overall process is illustrated by Figure 11: which shows the flow of information and processing in the self-learning diagnostic system.

It is apparent that the Historical database must be initialised, and preliminary data mining conducted before clinical application of the diagnostic system. The database will be initialised using a training set comprising data from animals with known metabolic conditions. Appropriate experimental design is vital to the construction of the initial

C:\WD0079\ProDoc\17A\Desklog\Documents\lgcontrol_as reviewed_final.doc-1471703

- 46 -

training data set. Empirical predictors derived using data mining are susceptible to artefactual relationships, involving nuisance factors – such as regional differences in diet and husbandry. For this reason, the training data set must be obtained from a multicentre trial, and stratified appropriately.

5

Specific Examples

Specific examples are set out in more detail in Appendix A. These are for the purpose of demonstration only and are not considered to be limiting.

- 10 Persons skilled in the art will appreciate that numerous variations and modifications will become apparent. All such variations and modifications which become apparent to persons skilled in the art, should be considered to fall within the spirit and scope that the invention broadly appearing before described.
- 15 Thus, for example, the above description has focussed on the testing of a general subject. It will be appreciated that this is most advantageously used for performance animals to identify conditions that may lead to a decrease in performance. This allows trainers to identify problems with horses or other animals before they would be noticeable using existing techniques. This is of particular benefit in the horse racing industry as it allows
- 20 problems to be identified in advance, which can in turn allow the conditions to be corrected before they effect the horses performance, which in turn can result in a vast loss of earnings for the trainers and owners of the horse.

However, the technique may also be applied to any subjects, including humans.

25

It will be appreciated that different predetermined data will be required for each type of subject being assessed.

C:\WD\DT\Project\VP\AD\Develop\Unassess\Specimen_01\rev1.doc 34/11/02

- 49 -

algorithms such as K-means, principal component analysis (PCA) and self-organising maps, commonly available in packages provided by companies such as spotfire, silicon genetics, and at higher levels of interpretation, Omniviz. These clusters amount to identified correlations 594 between gene expression and sample information and are stored in various formats, in the correlative database 595. An heuristic or neural network or rule-based computer software system pre-programmed with rules or training sets takes queries 534 (eg. expression details and sample details), stores these details in the sample database 593 and then compares the query pattern to those already stored in the correlative database 595 and produces standardized reports and correlation details 570 (according to the rules of the heuristic program). Correlation details are converted to useful information such as gene expression correlation results, for example a fully formatted report to include interpretations 571 and interpretations 575 (and optionally genes lists and gene sequences) and are securely delivered back to the requestor via the internet to Class One and Two users 505, 510.

Financials database 597 keeps track of details including for example accounting, purchasing and payroll details. Sales and marketing database 596 keeps track of items such as sales and marketing details, client details, customer relations management and stock management. Internal data warehouse 560 receives information from databases 590, 596 and 597. This internal data warehouse 560 will only be accessed by authorized internal users conducting legitimate business activities. A secure (internal) data warehouse 545 services the needs of Class Three users 515. Specific (and confidential) information 580 is extracted from internal data warehouse 560 that is then stored in secure customer data warehouse 545 where authorized users 515 can query 535 (for example as interrogation requests), specific and confidential information such as clinical history information, pathology results and interpretations. This information is presented in a secure user-friendly and/or visual format 585 in relation to individuals or groups of athletes or performance animals, and/or time series of results.

Figure 13 is a flow diagram of one specific example showing steps for assessing a biological sample for diagnosing or assessing a condition of an animal. A user collects a

C:\PENTUP\Bin\VPAT\NCC0000\ccr\alg\mcc02_05\app\Draw_Eml\Draw_14/11/02

- 52 -

processor 1140, which may be for example, a laptop computer with a modem. The formatted data is transmitted via a communications network 1150, for example the Internet. A remote diagnostic server 1160 receives the transmitted data and the data is compared with a database(s) 1161 which stores data, for example, data in relation to
5 nucleic acid location on an array, expression level (relative abundance or absolute abundance) of a nucleic acid hybridised with a corresponding nucleic acid on an array, and data correlating nucleic acid expression level and performance, health, or condition of an animal.

10 Figure 15 is a flow diagram illustrating steps for preparing an array. A biological sample 1210 is collected from an animal. Biological sample 1210 may comprise for example, a blood sample (preferably white blood cells isolated therefrom), urine sample or tissue sample (including fetal tissues and tissues in various stages of development). A specific aim of collecting the biological sample is to isolate and sequence as many relevant genes
15 from the sample for use on an array. Thousands of nucleic acids may be isolated that may form a large number of probes for a broad screening of an animal's genetic make-up or gene expression pattern.

Nucleic acids are isolated from the biological sample. In one instance the sample may be
20 used to prepare genomic DNA or tissue specific mRNA 1223. In another instance RNA is isolated from the biological sample 1210 and a cDNA library 1220 is prepared from the isolated RNA. Plasmids 1221 comprising cDNA inserts from library 1220 may be sequenced 1222 from either or both 5' and/or 3' end of the nucleic acid. Preferably, sequencing is from the 3' end. Sequences may comprise Expressed Sequence Tags (EST).
25 If an isolated nucleic acid does not encode a full-length gene (eg. an EST), a partial nucleic acid may be used as a probe to isolate a full-length nucleic acid. Alternatively, or in addition, EST sequence information may be compared directly with a sequence database 1230, for example GenBank, and a search for related or identical sequences performed. Putative gene identification and function 1231 may be determined from a search, for
30 example a BLAST search performed in step 1230. By determining the number of times

C:\WP007\Project\17A\Project\Documents\17A\17A_001.doc 14/11/02

- 53 -

each gene is represented in the library, a computer may be programmed to enable the normalisation and standardisation of the relative abundance data of mRNAs in a sample.

Gene-specific oligonucleotides 1232 may be synthesised using information from EST or full-nucleotide sequence 1222 data. Gene-specific oligonucleotides 1232 may be used as amplification primers to amplify (step 1224) a region of a corresponding nucleic acid. The nucleic acid used as template to amplify a region of corresponding nucleic acid may be, for example, isolated plasmid DNA 1221 and/or genomic DNA, cDNA or mRNA (eg. used with RT-PCR) 1223. The nucleic acid thus prepared can be used directly as the nucleic acids for attaching to an array 1240. Amplification products 1225 may also be generated using non-gene-specific primers (eg. oligo-dT, plasmid sequence flanking a nucleic acid of interest). Oligonucleotides corresponding to a gene 1232 may also be used on array 1240, alternatively the oligonucleotide corresponding to known sequence can be built successively nucleotide by nucleotide on a support using Affymetrix methodology such as that in US patent no. 5,831,070, incorporated herein by reference.

In one embodiment, the step relating to constructing cDNA 1220 and isolating plasmids 1221 comprising the cDNA may be omitted. In this embodiment, isolated genomic DNA or tissue specific mRNA 1223 is used as a template to make amplification product 1225 by amplification using gene-specific primers 1232. Amplification product 1225 may be attached to array 1240.

Nucleic acids attached to or built onto array 1240 preferably represent most, more preferably all, expressed genes in a given tissue from an animal of interest. For example, for a complete diagnostic test for racehorse blood, the array should contain genes expressed in the cells of blood under various conditions and at various stages of cell differentiation.

Figure 16 shows a flow diagram comprising steps for determining gene expression in biological samples comprising both reference target 1305 and sample target 1310. Nucleic acids, in particular RNA (total RNA or mRNA), are isolated from biological samples 1305

C:\WORK\77\ProG\1\VPAD\Develop\Documents\proG\1\proG_1441\1441.U02

- 55 -

unique and standard identifier. An example of such a system is currently used in clinical medicine and veterinary science and termed SNOMED or SNOVET (Standardised Nomenclature of Medicine or Veterinary Science), where a clinical condition can be described using a numerical system. This system has not been used for describing the normal condition or the ability of a performance animal to perform to its best. A numerical grading system could also be used to standardise the collection of such data, for example, time spent on a treadmill is a strong indicator of exercise tolerance, as is blood concentration of oxygen and ability to transport oxygen. Conditions may include disease, response to drugs, training, nutrition and environment. The clinical information 1460 is formatted into electronic form 1440, for example a digital signal, suitable for transmission via a communications network 1450.

The process is repeated such that a collection of several array readouts for particular conditions are made. A standard range (for example, a population median of 95%) of values for each of the represented genes and its relative abundance can be calculated. This reference range can then be used as a comparison to test sample results.

Nucleic acid expression information from a read array 1430 for a target sample is correlated with previously measured conditions 1460 to provide information on nucleic acid expression level (abundance or relative abundance) with any previously measured condition. This information is compiled at server 1470 and good data is stored and bad data rejected 1480. The compilation process includes collection of a large enough set of array readout information for a particular condition so that inferences can be drawn on gene expression profiles and conditions. The compilation 1470 may also include use of sophisticated pattern recognition and organisational software and algorithms (examples common to the art include algorithms such as K means, Anova and Mann Whitney, Self Organising Maps, principal component analysis, hierarchical clustering – any one of which is available as part of proprietary software packages) such that expression patterns that differ to normal or expected condition can be identified. The compilation 1470 will preferably include sophisticated methods of supervised classification such as regularised discriminant analysis, diagonal discriminant analysis, support vector machines, or

C:\WORK\11\Fred\DAVIES COLLISON CAVE Pat. & Trad\14-11-02

- 56 -

recursive partitioning – any one of which is readily conducted using proprietary software packages. Concurrently, comprehensive clinical information 1460 for animals may be collected and biological samples 1410 tested on arrays so that correlations can be made between any clinical observation and array data. In this manner a database is created
5 comprising data on nucleic acid expression which may include data correlating any desired condition, for example normal and specific abnormal condition(s), with nucleic acid expression. The stored data 1480 may be accessed using specific programs and algorithms
1490.

10 Throughout this specification, unless the context requires otherwise, the words comprise, comprises and comprising will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

In order that the techniques outlined above may be readily understood and put into
15 practical effect, particular preferred embodiments will now be described by way of the following non-limiting examples.

STEP 1

Biological Sample Collection

20 A biological sample comprising nucleic acids, for example total RNA and mRNA, is collected. The biological sample may include cells of the immune system at various stages of development, differentiation and activity. The biological sample in most instances would be whole blood collected from a vein of a performance animal. However, the biological sample may include a fluid and/or tissue, for example sputum, urine, tissue
25 biopsies, bronchial or nasal lavages, joint fluid, peritoneal fluid or thoracic fluid which, in part, comprises cells of the immune system that have infiltrated such tissues or fluids. Cells present in blood which comprise mRNA may include mature, immature and developing neutrophils, lymphocytes, monocytes, reticulocytes, basophils, eosinophils, macrophages. All of these cell types also appear in tissues of non-blood origin at various
30 times in various conditions.

CAPSOTV\Fred\VPAL\m\00p\000021\Agentic_k_00\revindreg_Card.doc 14/11/02

- 57 -

Methods described herein may include use of the abovementioned cell types. The biological sample is collected and prepared using various methods. For example, an easy method of collecting cells of the blood is by venipuncture. The biological sample may be collected from a performance animal, for example, a horse with suspected laminitis, a human athlete or camel with osteochondrosis, or a greyhound with subclinical cystitis.

Blood sample

Ten ml of blood is drawn slowly (to prevent hemolysis) from the vein of an animal (jugular vein in a horse and camel, veins on the forearm/limb of humans and dogs) into a 1:16 volume of 4% sodium citrate to prevent clotting and the sample is mixed and then placed on ice. The sample is centrifuged at 3000 RPM at 40C for 15 minutes and white blood cells (WBC) (commonly called the "buffy coat") are removed from the interface between plasma and red blood cells (RBC) into a separate tube using a pipette. The WBCs are then treated with at least 20 volumes of 0.8% ammonium chloride solution to lyse any contaminating RBC and re-centrifuged at 3000 RPM at 40C for 5 minutes. The pelleted WBCs are then washed in 0.9% sodium chloride, re-centrifuged, and kept on ice. The cell pellet is then used directly in RNA extraction.

Non-blood biological fluid sample

A fluid sample, for example, sputum, urine, bronchial or nasal lavages, joint fluid, peritoneal fluid or thoracic fluid, is centrifuged at 3000 RPM at 4 0C for 20 minutes to collect cells. Samples comprising large amounts of mucous are treated with a mucolytic agent such as dithiothreitol prior to centrifugation. A cell pellet is then washed in 0.9% sodium chloride, re-centrifuged and the cell pellet is used directly in RNA extraction.

Tissue biopsy

A tissue biopsy is frozen in dry ice or liquid nitrogen and crushed to powder using a mortar and pestle. The frozen tissue is then used directly in RNA extraction.

STEP 2

RNA Isolation

CA\WIP\NT\ProDev\VP\AID\topol\Gencard\gencard_00_revision_1.mtl.doc 14/11/02

- 58 -

Total RNA and/or mRNA is isolated from a biological sample. Use of isolated mRNA rather than total RNA may provide results with less background and improved signal.

RNA is commonly isolated by skilled persons in the art, and examples of some methods
5 for isolating mRNA are described below.

Commercially available kits, for example, Qiagen RNA and Direct RNA extraction kits, and RNA extraction kits produced by Invitrogen (formerly Life Technologies) and Amersham Pharmacia Biotech herein incorporated by reference, may be used by following
10 the manufacturer's instructions. Key elements of these mRNA extraction protocols include use of an appropriate amount of sample, protection of the sample from RNase contamination, elution of the sample from a column at 70°C and quantitation and quality checking in an agarose 0.7% gel and using an OD 260/280 ratio. About 0.2 gm (wet weight) of pelleted white blood cells or tissue is required for each mRNA extraction which
15 will yield about 1-2 µg of mRNA. Disposable gloves should be worn throughout the procedure, with frequent changes. Both the column and solution used for elution should be at 70°C.

RNA quantification and assessment of RNA size and quality include standard gel
20 electrophoresis methods of running a small quantity of an RNA sample on an agarose gel with known standards, staining the gel with for example ethidium bromide to detect the sample and standards and comparing relative intensities and size of standard RNA and sample RNAs, comparison of the intensities of the ribosomal RNA bands. Alternatively, or in addition, RNA concentration in a solution may be determined by measuring
25 absorbance at 260/280 nm in a spectrophotometer relative to known standards and calculated using known formulas.

cDNA Synthesis and Labelling

RNA prepared as described above may be synthesised to cDNA and labelled resulting in a
30 labelled probe using kits provided by suppliers such as Amersham Pharmacia Biotech, Invitrogen, Stratagene or NEN, herein incorporated by reference. For example, a typical

- 60 -

A nucleic acid may be labelled with one or more labelling moieties for detection of hybridised labelled nucleic acid (ie. probe) and target nucleic acid complexes. Labelling moieties may include compositions that can be detected by spectroscopic, photochemical, biochemical, immunochemical, optical or chemical means. Labelling moieties may include radioisotopes, such as ^{32}P , ^{33}P or ^{35}S , chemiluminescent compounds, labelled binding proteins, heavy metal atoms, spectroscopic markers, such as fluorescent markers and dyes, magnetic labels, linked enzymes, and the like. Preferred fluorescent markers include Cy3 and Cy5, for example available from Amersham Pharmacia Biotech (as described above).

25 The Affymetrix system uses RNA as substrate and generates biotin labelled cRNA through a series of reactions detailed in a protocol available from their website (affymetrix.com), incorporated herein by reference. The cRNA is fragmented prior to application onto the array.

STEP 3

Arrays

C:\WP\TVP\Films\VFA\Unidoc\Genes\dog\grain23_01 rev1.doc, the 11/02

- 61 -

One feature is an array comprising nucleic acids representing expressed genes from cells found in blood of a performance animal, for example a horse, human, camel or dog. The nucleic acids may be of any length, for example a polynucleotide or oligonucleotide as defined herein.

5

Each nucleic acid occupies a known location on an array. A nucleic acid target sample probe is hybridised with the array of nucleic acids and an amount or relative abundance of target nucleic acid hybridised to each probe in the array is determined.

- 10 High-density arrays are useful for monitoring gene expression and presence of allelic markers which may be associated with disease. Fabrication and use of high density arrays in monitoring gene expression have been previously described, for example in WO 97/10365, WO 92/10588 and US Patent No. 5,677,195, all incorporated herein by reference. In some embodiments, high-density oligonucleotide arrays are synthesised
- 15 using methods such as the Very Large Scale Immobilised Polymer Synthesis (VLSIPS) described in US Patent No. 5,445,934, incorporated herein by reference.

- Arrays for human are commercially available from companies such as Incyte, Research Genetics, and Affymetrix. Lion Bioscience recently announced forthcoming release of a
- 20 dog microarray and have a clone collection of dog cDNAs. These arrays typically comprise between 2,000 and 10,000 genes and are species specific. None are available for the horse or camel. Some of these genes are in multiple copies on the array and have not been fully annotated or given a true gene identity. Additionally, it is not known whether DNA on the array, when hybridised to a test sample, specifically binds to a single gene.
- 25 This latter instance results from splice variants of RNA transcripts in tissues such that one gene may encode multiple transcripts.

- Human and dog arrays (when available) can be used in methods described herein. However, these arrays are currently non-specific and include genes that are not expressed
- 30 in blood cells of animals, and/or do not contain genes important in controlling the function of blood cells, and/or contain regions of genes that are not specific to blood cells.

C:\WD\NTV\work\CV\A\Destiny\General\Agreement_no_revision_1471102

- 62 -

Clones containing specific genes are available and can be purchased for human (mouse and dog) for use on arrays (for example from the IMAGE consortium or Lion Bioscience). However, it is not possible to obtain specific clones for use on a blood-specific array without prior knowledge of what genes are expressed in blood cells. The IMAGE consortium also does not guarantee that the gene of interest is contained in the clone purchased.

Array Construction

Because of difficulties, problems and a likelihood of wasting financial resources to obtain a blood-specific DNA array, a method is provided herein which provides rapid and cost effective generation of species and tissue-specific DNA arrays for assessing nucleic acid expression in a sample. Figure 14 shows steps for constructing an array in one embodiment.

Target Nucleic Acid Preparation

Biological samples are collected as described above. Samples comprising cells expressing as many genes of interest in relation to condition(s) of a performance animal are collected. For example, a sample comprising a mixture of nucleated blood cells from performance animals with conditions such as, osteochondrosis, laminitis, tendon soreness, bursitis, abscesses, inflammation, allergy, viral infection, parasite infection, asthma, etc.

Approximately 5 μ g of mRNA is isolated from the biological sample (typically 1 gm wet weight) using mRNA isolation kits or the protocol described above. Concurrently, 5 μ g of mRNA is isolated from umbilical cord blood, and/or early stage foetus. Cells and tissues contained within these sources would express genes that may not be expressed in the cells extracted from blood in the above example. Isolation of cytoplasmic mRNA from cells is preferred. This step involves rupturing the cells with a solution comprising detergent and/or chaotropic agent and salt such that cell nuclei and the nuclear membrane remain intact. The cell nuclei are pelleted by centrifugation and the supernatant is used for mRNA extraction. Protocols for this procedure are available as part of mRNA isolation kits (eg

25 Subjective factors influencing whether a nucleic acid should be used on an array include
quality and confidence of the DNA sequence, a Genbank homology score with identified
nucleic acids, evidence of a poly-A tail (indicative of a translated transcript), uniqueness of
the 3' sequence data (compared to both Genbank and an in-house database of clone
30 sequences).

- 64 -

Nucleotide sequences may be compared with an existing database, for example Genbank, to determine a previously provided name, tissue expression, timing of expression, biochemical pathway, cluster membership, and possible function or cellular role of an expressed nucleic acid. In addition, a nucleic acid fragment may be used as a probe to isolate a full-length nucleic acid which may encode a gene which is associated with a particular disease or condition. Further, identified nucleic acids may be used to isolate homologues thereof, inclusive of orthologues from other species. An identified nucleic acid may also be cloned into a suitable expression vector to produce an expressed polypeptide in vitro, which may be used, for example as an antigen in generating antibodies and for use on protein arrays. The antibodies may be used for developing specific diagnostic assays or therapies, for three-dimensional protein structure such as X-ray crystallographic studies, or for therapeutic development.

C:\WDEV\Pro\Q\VP\A\Develop\Genes\B\genet11_on_nov1998_02a1.doc 14/11/02

- 65 -

An array may comprise any number of different nucleic acids, but typically comprises greater than about 100, preferably greater than about 1,000, more preferably greater than about 5,000 different nucleic acids. An array may comprise more than 1,000,000 different nucleic acids. Each nucleic acid is preferably represented more than once for scanning
5 internal comparison and control. Preferably, the nucleic acids are provided in small quantities and are gene-specific and/or species-specific usually between 50 and 600 nucleotides long, arranged on a solid support.

The Affymetrix system uses 11 probes per gene, each of 25 nucleotides, that are built onto
10 the array using a photolithographic method (US Patent Nos. 6,309,831; 6,168,948; 5,856,174; 5,599,695; 5,831,070; 6,153,743; 6,239,273; 6,271,957; 6,329,143; 6,310,189 and 6,346,413). The nucleic acids may be dotted onto the solid support or bound to microspheres, or in solution. A typical array may have a surface area of less than 1 cm², for example a microarray.

15 A nucleic acid can be attached to a solid support via chemical bonding. Furthermore, the nucleic acid does not have to be directly bound to the solid support, but rather can be bound to the solid support through a linker group. The linker groups may be of sufficient length to provide exposure to the attached nucleic acid. Linker groups may include
20 ethylene glycol oligomers, diamines, diacids and the like. Reactive groups on the solid support surface may react with one of the terminal portions of the linker to bind the linker to the solid support. Another terminal portion of the linker is then functionalised for binding the nucleic acid. A solid support may be any suitable rigid or semi-rigid support, including charged nylon or nitrocellulose, chemically treated glass slides available from
25 companies such as NEN, Corning, S&S, arrays available through Affymetrix, membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The solid support can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which the nucleic acids are bound. Preferably, the solid support is optically transparent.

30

- 66 -

An example for preparing an array element is provided by the manganese superoxide dismutase gene. A clone comprising a nucleic acid insert is prepared and isolated as described above. The clone is sequenced to identify the nucleotide sequence. A BLAST search using the identified nucleotide sequence is performed to determine homology of the cloned nucleic acid with nucleic acids in a database, for example GenBank. Identification of nucleotide sequence homology with superoxide dismutase genes stored in the database provides a level of confidence that the clone comprises at least in part a gene for superoxide dismutase for the horse. Unique primers can be designed to amplify a nucleic acid using PCR and the clone DNA, or genomic DNA from the same species as a template. Purified amplification product can be directly attached to an array and thereby act as a target for a complementary labelled nucleic acid probe in the test and reference samples. Alternatively, a unique sequence can be determined and an oligonucleotide manufactured

C:\W200TV\620\177\A\Doc\p4\Gen22\p4\Gen22_10_0001.doc 14/11/02

- 68 -

for superoxide dismutase may be the only gene up-regulated in a particular condition, especially in conditions of inflammation, or a large number of genes may be up- and down- regulated in various conditions. It is this fingerprint, rather than specific knowledge of gene sequence or function that can be used as a marker for various conditions. It would
5 be expected that fingerprints be useful across species barriers to include performance animals such as humans, horse, dog and camel.

The arrangement of nucleic acids on the array may be periodically changed and these arrays are then assigned a particular batch code that corresponds to a specific array
10 comprising a specific nucleic acid arrangement. The ability to change the arrangement of nucleic acids on the array and knowledge of the exact arrangement may prevent other people from generating a database using the arrays described above. Using a batch code also enables tracking of manufacturers of the arrays in regards to the number of arrays produced. The batch code further enables validation of a user of the communication
15 network or "internet" diagnostic method and system. Batch code can also identify a particular type of array used, should more disease-specific arrays be designed and manufactured.

An example of how an array may be prepared and analysed is described in Eisen and
20 Brown (Methods in Enzymology, 1999, 303 179) and in US Patent No. 6,114,114, herein incorporated by reference. Chapter 22 of Ausubel et al. supra also describes methods and apparatus for use with arrays and is herein incorporated by reference.

Control samples may be respectively labelled in parallel with a test and reference sample.
25 Quantitation controls within a sample may be used to assure that amplification and labelling procedures do not change a true distribution of nucleic acid probes in a sample. For this purpose, a sample may include or be "spiked" with a known amount of a control nucleic acid which specifically hybridises with a control target nucleic acid. After hybridisation and processing, a hybridisation signal obtained should reflect accurately
30 amounts of control nucleic acid added to the sample. For such purposes, a microarray may have internal controls, for example a nucleic acid encoding a common gene expressed by

C:\W\001\ProDev\AID\Doc\log\control_log_revision_14/11/02

- 69 -

the performance animal with known expression levels and a nucleic acid encoding a gene from another species that is known not to hybridise to the test or reference sample. To improve sensitivity and specificity of the assay, blocking agents such as Cot DNA from the tested species may also be used.

5

STEP 4

Hybridising Sample Nucleic Acid Probes with an Array

Nucleic acid probes may be prepared as described above from a biological sample from a performance animal that has been assessed concurrently by physical inspection and/or blood tests or other method. Nucleic acid targets from a statistically relevant number of normal animals previously hybridised to arrays, and a reference range for each of the genes on the array is calculated and used as a normal reference range (for example a 95% population median). Results from a test sample from a test animal can be compared with the same genes as the normal reference to determine if the test sample falls within the normal reference range. Further, nucleic acid targets may also be prepared from biological samples from apparently normal animals, animals with overt disease, various progressive stages of disease, hitherto undiagnosed or unclassified conditions or stages of such conditions, animals treated with known amounts of drugs (legal or otherwise), animals suspected of being treated with drugs (legal or otherwise), animals under specific exercise regimes for the sake of performance, animals subjected to (intentional or not) various nutritional states and/or environmental conditions. Databases of information from the use of such samples and arrays are created such that test samples can be compared. The database will then contain specific patterns of gene expression for particular conditions.

25

Prior to hybridisation, a nucleic acid probe may be fragmented. Fragmentation may improve hybridisation by minimising secondary structure and/or cross-hybridisation with another nucleic acid probe in a sample or a nucleic acid comprising non-complementary sequence. Fragmentation can be performed by mechanical or chemical means common in the art.

30

C:\P\DOT\Probes\VP\A\array\Control\array2_10_revision_1.doc 14/11/02

- 70 -

A labelled nucleic acid target may hybridise with a complementary nucleic acid probe located on an array. Incubation conditions may be adjusted, for example incubation time, temperature and ionic strength of buffer, so that hybridisation occurs with precise complementary matches (high stringency conditions) or with various degrees of less complementarity (low or medium stringency conditions). High stringency conditions may be used to reduce background or non-specific binding. Specific hybridisation solutions and hybridisation apparatus are available commercially by, for example, Stratagene, Clontech, Geneworks.

- 10 Affymetrix have detailed a standard procedure for the hybridisation of probes with an array (as describe at their website, affymetrix.com, incorporated herein by reference), however, a typical method entails the following:

- Adjust probe volume (prepared as above) to a value indicated in the "Probe & TE" column below according to the size of the cover slip to be used and then add the appropriate volume of 20XSSC and 10% SDS.

Cover Slip Size (mm)	Total Hyb Volume (l)	Probe & TE (l)	20x SSC (l)	10% SDS (l)
22 x 22	15	12	2.55	0.45
22 x 40	25	20	4.25	0.75
22 x 60	35	28	5.95	1.05

20xSSC is 3.0 M NaCl, 300 mM NaCitrate (pH 7.0).

20

- Denature the probe by heating it for 2 min at 100°C, and centrifuge at 14,000 RPM for 15-20 min. Place the entire probe volume on the array under the appropriately sized glass cover slip. Hybridize at 65°C (temperatures may vary when using different hybridisation solutions) for 14 to 18 hours in a custom slide chamber (for example a Corning CMT hybridisation chamber #2551).

25

C:\W2\NT\ProGen\VPADesktop\userdata\general\as\retail\as_001.doc-14/11/02

- 71 -

Washing the Array

After hybridisation, the array is washed to remove non-specific probe and dye hybridisation. Wash solutions generally comprise salt and detergent in water and are commercially available. The wash solutions are applied to the array at a predetermined

5 temperature and can be performed in a commercially available apparatus. Stringency conditions of the wash solution may vary, for example from low to high stringency as herein described. Washing at higher stringency may reduce background or non-specific hybridisation. It is understood that standardisation of this step is required to produce maximum signal to noise ratio by varying the concentration of salt used, whether detergent

10 is present (SDS), the temperature of the wash solution and the time spent in the wash solution.

A typical wash protocol consists of removing the slide from a slide chamber, removing the cover slip and placing the slide into 0.1%SSC (recipe provided above) and 0.1% SDS at

15 room temperature for 5 minutes. Transfer the slide to 0.1% SSC for 5 minutes and repeat. Dry the slide using centrifugation or a stream of air. Equipment is available to enable the handling of more than one slide at a time (for example, slide racks).

STEP 5

20 Reading the Array

After removal of non-hybridised probe, a scanner or "array reader" is used to determine the levels and patterns of fluorescence from hybridised probes. The scanned images are examined to determine degree of hybridisation and the relative abundance of each nucleic acid on the array. A test sample signal corresponds with relative abundance of an RNA

25 transcript, or gene expression, in a biological sample. Alternatively, an Affymetrix array is read and computer algorithms calculate the difference between hybridisation on perfect match and mismatch probes for each of the 11 probes sets for each gene. It then calculates a presence or absence, an absolute value for each gene and a p value for the absolute call.

30 Array readers are available commercially from companies such as Axon and Molecular Dynamics and Affymetrix. These machines typically use lasers, and may use lasers at

C:\WINNT\Profiles\VP\Documents\Control\grades1_no revision_Guid.doc-14/11/02

- 73 -

- b) Integument: eyes, sores, abscesses, wounds, insects/parasites, allergy, infection.
- c) Cardio/Respiratory: eyes, nasal discharge, rales, viral/bacterial infection, allergy, chronic obstructive pulmonary disease, cough/wheeze, crepitous sounds in the thorax, epistaxis, auscultation sounds, heart sounds, capillary refill, mucous membrane colour.
- 5 d) Gastrointestinal: diarrhoea, colic/stasis, parasites, appetite level, drenching time and dose.
- e) Reproductive: stage of pregnancy, abortion, inflammation, discharges.
- f) Musculoskeletal: lameness, laminitis, bone or shin soreness, muscle soreness or tying up, tendon or ligament affected, level of pain, X-ray data, scintigraphy data, CAT scan
- 10 data, bursitis, bruising, cramping or "tying up".
- g) Blood test results: biochemistry, immunology, serology (viral, bacteriological, hormone levels), cell counts, cell morphology, pathologist interpretation.
- h) Other diagnostic test results: X-ray, biopsy, histopathology, CAT scan, MRI, bacteriology, virology.
- 15 i) Other data: Season (date), location, male or female, vaccination history, body score (fitness and fat), fitness level.

Alternatively, the entire system could be based on the aforementioned SNOMED system with appropriate modifications to encompass descriptions of exercise physiology and the

20 normal animal. Alternatively, the entire system could rely on text or categorical data that can be appraised and scored by software such as Omniviz. Whatever system is used, it would be appreciated that the aim is to adequately, systematically and in a standard manner describe the current condition of the animal to the best of currently available technologies and could include results from machinery such as X-ray, ultrasound, scintigraphy and

25 blood analysis.

The user also ensures that array results (that may for example be automatically collected from a reader), array specifications, data mining specifications, level of interpretation required and the clinical information are entered and correspond to the same animal and

30 the same sample. The form is transmitted electronically to a central database and recognised as an individual accession or request by the database. The central database

- 74 -

The above steps may be automated so that a user need not be present to perform the tasks. In an automated specific example, gene expression data from an array reader may be transmitted via a communications network directly to a server which is connected to a central database. Additional information could be input by the user at a processor which is also linked to the array reader.

15 A central database interprets the array specifications (eg. nucleic acid order on a
microarray), decodes the information transmitted, determines nucleic acid expression level
in a biological sample and compares the expression level and patterns of expression with
known standards or reference range. Various levels of database interpretation may be
applied to the data transmitted, depending on the user requirements. Clusters of genes may
be up-regulated or down-regulated in certain conditions and the database makes automated
20 correlations to specific conditions by accessing various levels of database information.

Mining software such as Metamine (Silicon Genetics), ArraySCOUT (Lion Bioscience) can be used in this instance, and more advanced data mining technologies could be used to identify patterns and nearest neighbour information in data (such as products from AnVil Informatics Inc and OmniViz Inc). Further, software capable of taking rule-based instructions (such as that described by Pacific Knowledge Systems Sydney Australia in their "ripple down" technology) and having the ability to self learn (heuristics and neural network systems) such as that described in Khan et al. Nature Medicine 7 (6) 673, incorporated herein by reference, could be used at this stage to limit the level of human interaction in determining a diagnosis. In this latter example, an artificial neural network is used, and samples are divided into training and validation sets to create trained

C:\PDP\T\G\G\VPAD\Doc\app\Documents\general\as revisions_Final.doc-14/11/02

- 75 -

calibrated models. The calibrated models are then used to rank genes in diagnostic importance.

Levels of database may include:

- 5
 - Unique gene sequences (eg 3' and 5' EST sequence of genes)
 - Gene identity, homologous genes, tissue expression, keywords, function, cellular role, gene clusters, biochemical pathway, PubMed references
 - Primer sequences used to generate amplification products (eg two primer sequences used to uniquely amplify the gene for gamma interferon in a particular species)
- 10
 - Microarray construction and format (eg coded information on array manufacture batch and identification of genes and position on the array)
 - Blood profile and clinical data associated with particular conditions (eg standard clinical information and IDEXX-machine generated blood profile data)
 - Array data for normal and apparently normal status (eg 95% median range for
- 15
 - normal animals)
 - Array data for inducible disease and disease models
 - Array data for various overt diseases (eg joint inflammation)
 - Array data for stages of various overt diseases (eg pre-clinical, clinical and recovery stages)
- 20
 - Array data for the influence of various classes of drugs, legal or otherwise, of known administration and dose, or unknown administration or dose (eg various steroids)
 - Array data for the response to known and various levels of drugs used as a therapy (eg various anti-inflammatory medication at specific doses for a specific condition)
- 25
 - Array data for the response to exercise and various training regimes
 - Array data for the response to nutrition and various feeding regimes
 - Array data for the response to the environment so as to possibly determine influence of during various seasons, or allergens or feed types.
- 30 Each successive level relies on at least one previous level of database to allow for interpretation. The database may be built over time and more intensive searching of the

C:\WP51\PROG\01\VPAD\Doc\Exp\Oncology\genetics_rev\rev1.doc 14/11/02

- 76 -

database may incur a greater cost. As the database grows, changes may be made to the above methodology to increase the sensitivity of the detection of variation in expression of condition-specific genes – this could include the use of condition-specific arrays or condition-specific primers. Condition-specific arrays can be manufactured by a company
5 such as Affymetrix (under instructions) that would allow for increased sensitivity and specificity, much reduced size of arrays, decreased cost of production, and the ability to process multiple samples at once. The process of building the database is iterative, such that specific genes are correlated to specific conditions, and the detection of variations in these genes becomes more sensitive and specific through the use of various modifying
10 processes through the procedure (eg. the use of gene-specific primers for the amplification and labelling of cDNA from RNA, and the selection of limited numbers of genes on a disease- or condition-specific array, detection of splice variants and single nucleotide polymorphisms).

15

STEP 7

Standardised Electronic Reporting

The database reports back electronically to a remote user, either automatically or with a level of human intervention. The electronic report may be converted to a printed document. The report provides details of an animal's condition that is determined by
20 correlation of gene expression data with information stored in a remote database, and optionally expert analysis.

Information sent might include:

- Individual genes up-regulated or down-regulated (for example, with laminitis or
25 joint capsule inflammation or bursitis, a report on the up-regulation of genes such as interleukin-3, manganese superoxide dismutase, Gro α , metalloproteinase matrix-metallo-elastase, ferritin light chain may have some correlation to tissue inflammation, and down-regulation of genes such as insulin-like growth factor and its receptor may be correlated to recovery from such a condition). The identity of these genes cannot be
30 predicted to be associated to any condition unless the above described methodology is used and databases on relative expression of genes for particular conditions have been

C:\PDP\T\Protein\VF\A\Doc\p4\Genetic\align\1_in\revision_1\test.doc 14/11/02

- 77 -

compiled. Therefore a screening test covering all genes may need to be performed first and a second, more specific test then applied.

• The overall pattern of gene expression and any correlation to particular conditions. For example, animals in heavy training may have a gene "fingerprint" that is different to animals being spelled from training.

• Individual pattern of gene expression (ie. the shape of the gene expression pattern over a time course or multiple samples taken over a period may change as an animal recovers from a condition)

• Changes to a pattern of gene expression, gene expression profile or level for a single animal over a time period or for successive tests.

• Clusters of genes up-regulated or down-regulated in a particular condition

• Pathways of genes up-regulated or down-regulated in a particular condition

• Correlations between genes up-regulated or down-regulated and known conditions, or stage of condition, or influence

• Known therapies to ameliorate the condition or enhance desired effects

• Specialist pathologist written interpretation

• Relevant information of use to veterinarians, medical practitioners, owners, trainers and athletes

• Collections of data on groups of animals under specific management regimes

Throughout the specification the aim has been to describe the preferred embodiments of the invention without limiting the invention to any one embodiment or specific collection of features. It would therefore be appreciated by those of skill in the art that, in light of the instant disclosure, various modifications and changes can be made in the particular embodiments exemplified without departing from the scope of the present invention. For example, the examples described herein may be used with performance animals other than horse, for example human, dog and camel.

All references, inclusive of patents, patent applications, scientific documents and computer programs, referred to in this specification are herein incorporated by reference in its entirety.

C:\W2001\VP\A\Davies\patent\genomics_rev\doc_02a.doc-14/11/02

- 78 -

Dated this of Fourteenth day of November, 2002

GENOMICS RESEARCH PARTNERS PTY LIMITED

5 By their Patent Attorneys

DAVIES COLLISON CAVE

- 79 -

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

- 1) A method of determining the status of a subject, the method including:
- a) Obtaining subject data, the subject data including one or more parameter values, at least one of the parameters being indicative of the current biological status of the subject;
- b) Comparing the subject data to predetermined data, the predetermined data including for each of a number of individuals:
- i) One or more values for the parameters for the respective individual, at least some of the individuals having a number of conditions relevant to the status of the individual, the number of parameters being statistically sufficient to distinguish each of the conditions; and,
- ii) An indication of the status of the respective individual; and,
- c) Determining the status of the subject in accordance with the results of the comparison.
- 2) A method according to claim 1, the number of parameters being greater than 100.
- 3) A method according to claim 1, the number of parameters being greater than 1000.
- 4) A method according to claim 1, the number of parameters being less than 6000.
- 5) A method according to claim 1, the method including determining any conditions displayed by the user.
- 6) A method according to claim 5, the method including determining the ability of the subject to perform in a sporting and/or racing event in accordance with any determined conditions.
- 7) A method according to claim 1, the method of performing the comparison including causing the second processing system to:
- a) Obtain a set of templates, the set of templates representing differences between groups of individuals; and,
- b) Use the templates to classify the subject data into a respective one of the groups.
- 8) A method according to claim 7, the method including determining one or more conditions in accordance with the determined groups.
- 9) A method according to claim 1, the parameters being representative of the level or abundance of a molecule selected from one or more of :

C:\WDNT\PreGo\177A\Doc\177A\Doc\177A\177A_141102

- 80 -

- a) A polynucleotide;
 - b) A polypeptide; and,
 - c) A polysaccharide.
- 10) A method according to claim 1, the method including:
- 5 a) Receiving confirmation of the determined status; and,
- b) Updating the predetermined data in accordance with the confirmed status and the subject data.
- 11) A method according to claim 1, the predetermined data including phenotypic information of the individuals, and the subject data including phenotypic information
- 10 regarding the subject, the phenotypic information including details of one or more phenotypic traits.
- 12) A method according to claim 11, the method including comparing the subject data to predetermined data for individuals having one or more phenotypic traits in common with the subject.
- 15 13) Apparatus for determining the status of a subject, the apparatus including a processing system adapted to:
- a) Obtain subject data, the subject data including one or more parameter values, at least one of the parameters being indicative of the current biological status of the subject;
- 20 b) Compare the subject data to predetermined data, the predetermined data including for each of a number of individuals:
- i) One or more values for the parameters for the respective individual, at least some of the individuals having a number of conditions relevant to the status of the individual, the number of parameters being statistically sufficient to
- 25 distinguish each of the conditions; and,
- ii) An indication of the status of the respective individual; and,
- c) Determine the status of the subject in accordance with the results of the comparison.
- 14) A computer program product for determining the status of a subject, the computer
- 30 program product including computer executable code which when executed on a

C:\WDONT\TheDaviesCollisonCavePat&Trad\revisions\Revised 14/11/02

- 81 -

suitable processing system causes the processing system to perform the method of claim 1.

15) A method of allowing a user to determine the status of a subject, the method including:

a) Receiving subject data from the user via a communications network, the subject data including one or more parameter values, at least one of the parameter being indicative of the current biological status of the subject;

b) Causing the base station to:

i) Compare the subject data to predetermined data, the predetermined data including for each of a number of individuals:

(1) One or more parameter values for the respective individual; and,

(2) An indication of the status of each individual; and,

ii) Determine the status of the subject in accordance with the results of the comparison; and,

c) Transferring an indication of the status of the subject to the user via the communications network.

16) A method according to claim 15, the method including:

a) Having the user determine the subject data using a remote end station; and,

b) Transferring the subject data from the end station to the base station via the communications network.

17) A method according to claim 15, the base station including first and second processing systems, the method including:

a) Transferring the subject data to the first processing system;

b) Transferring the subject data to the second processing system; and,

c) Causing the second processing system to perform the comparison.

18) A method according to claim 17, the method including:

a) Transferring the results of the comparison to the first processing system; and,

b) Causing the first processing system to determine the status of the subject.

19) A method according to claim 17, the method including at least one of:

a) Transferring the subject data between the communications network and the first processing system through a first firewall; and,

C:\WD\NT\Andr\VPADesktop\Genset\gpc\gpc_001\revisions_GenSet.doc-14/11/02

- 82 -

- b) Transferring the subject data between the first and the second processing systems through a second firewall.
- 20) A method according to claim 17, the second processing system being coupled to a database adapted to store the predetermined data, the method including:
- 5 a) Querying the database to obtain at least selected predetermined data from the database; and,
- b) Compare the selected predetermined data to the subject data.
- 21) A method according to claim 17, the second processing system being coupled to a subject database, the method including storing the subject data in the subject database.
- 10 22) A method according to claim 15, the status including details of any conditions of the individuals, the method including determining any conditions displayed by the user.
- 23) A method according to claim 22, the method including determining the ability of the subject to perform in a sporting and/or racing event in accordance with any determined conditions.
- 15 24) A method according to claim 22, the method of performing the comparison including causing the second processing system to:
- a) Obtain a set of templates, the set of templates representing differences between groups of individuals; and,
- b) Use the templates to classify the subject data into a respective one of the groups.
- 20 25) A method according to claim 24, the method including determining one or more conditions in accordance with the determined groups.
- 26) A method according to claim 15, the parameters being representative of the level or abundance of a molecule selected from one or more of:
- 25 a) A polynucleotide;
- b) A polypeptide; and,
- c) A polysaccharide.
- 27) A method according to claim 15, the method including having the user determine the subject data using a secure array, the secure array having a number of features each located at respective position on the array, and a respective serial number, the method
- 30 including causing the base station to:
- a) Determine the serial number from the subject data;

C:\PCT\ProG\VP\A\Doc\p\General\g\main_4a.pvt\438_1\001.doc-14/11/02

- 83 -

- b) Determine a layout indicating the position of each feature on the array;
 - c) Determining the parameter values in accordance with the determined layout, and the subject data.
- 28) A method according to claim 15, the method including:
- 5 a) Receiving confirmation of the determined status from the user; and,
 - b) Updating the predetermined data in accordance with the confirmed status and the subject data.
- 29) A method according to claim 27, the features including at least one of:
- 10 a) An oligonucleotide;
 - b) A peptide; and,
 - c) An antibody.
- 30) A method according to claim 15, the predetermined data including phenotypic information of the individuals, and the subject data including phenotypic information regarding the subject, the phenotypic information including details of one or more phenotypic traits.
- 15 31) A method according to claim 17, the method including comparing the subject data to predetermined data for individuals having one or more phenotypic traits in common with the subject.
- 32) A method according to claim 15, the method including causing the base station to:
- 20 a) Determine payment information, the payment information representing the provision of payment by the user; and,
 - b) Perform the comparison in response to the determination of the payment information.
- 33) A base station for determining the status of a subject, the base station including:
- 25 a) A store method for storing predetermined data, the predetermined data including for each of a number of individuals:
 - i) One or more parameter values, at least one of the parameters being indicative of the current biological status of the individual;
 - ii) An indication of the status of the individual; and,
 - 30 b) A processing system, the processing system being adapted to:

C:\VSOFT\ProDev\77A\Dev\77A\Gen\td\gen\td_77A\td_77A_14/11/02

- 87 -

- a) Obtain subject data, the subject data including one or more parameter values, at least one of the parameter being indicative of the current biological status of the subject;
- b) Compare the subject data to predetermined data, the predetermined data including for each of a number of individuals:
- 5 i) One or more parameter values for the respective individual; and,
- ii) An indication of the status of each individual;
- c) Determine the status of the subject in accordance with the results of the comparison; and,
- 10 d) Provide an indication of the ability in accordance with the results of the comparison.
- 49) A computer program product for determining the ability of a subject to perform in a sporting and/or racing event, the computer program product including computer executable code which when executed on a suitable processing system causes the processing system to perform the method of claim 46.
- 15 50) A method of providing secure arrays for use, each array including a number of predetermined features, the method including:
- a) Determining a number of respective feature layouts, each layout representing the positioning of each feature on a respective array;
- 20 b) Determining a number of serial numbers, each serial number corresponding to a respective layout;
- c) Generating a number of arrays, each array being generated in accordance with a respective layout, and including the corresponding serial number thereon, the serial number being used in processing used the array.
- 25 51) A method according to claim 50, the method being performed to provide the arrays on behalf of an entity, the method including providing an indication of the layouts and corresponding serial numbers to the entity, to thereby allow the entity to process the arrays.
- 30 52) A method according to claim 50, the method of determining the layouts including:
- a) Determining a preferred layout; and,

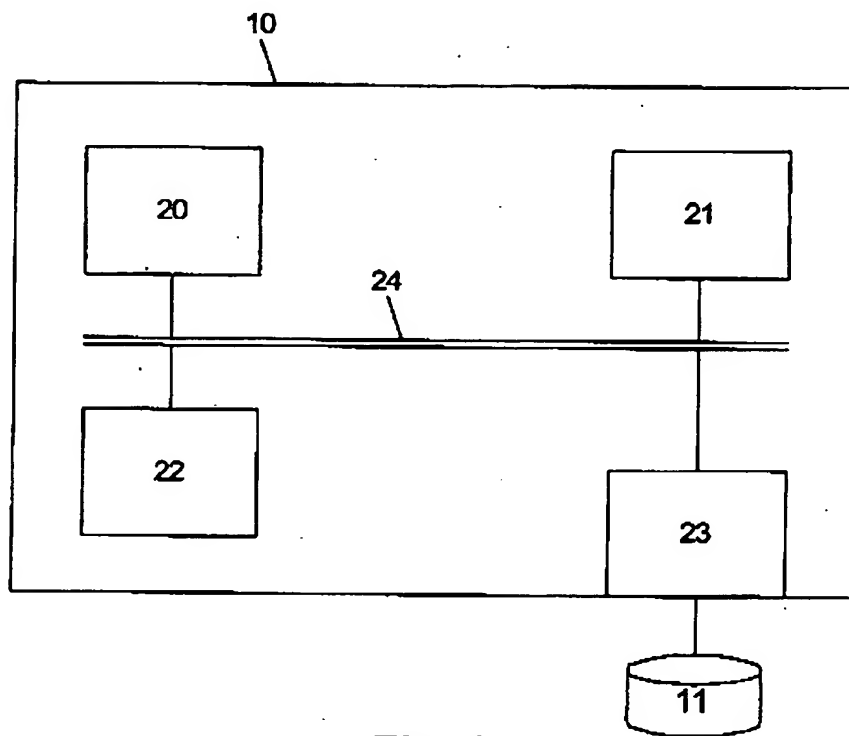


Fig. 1

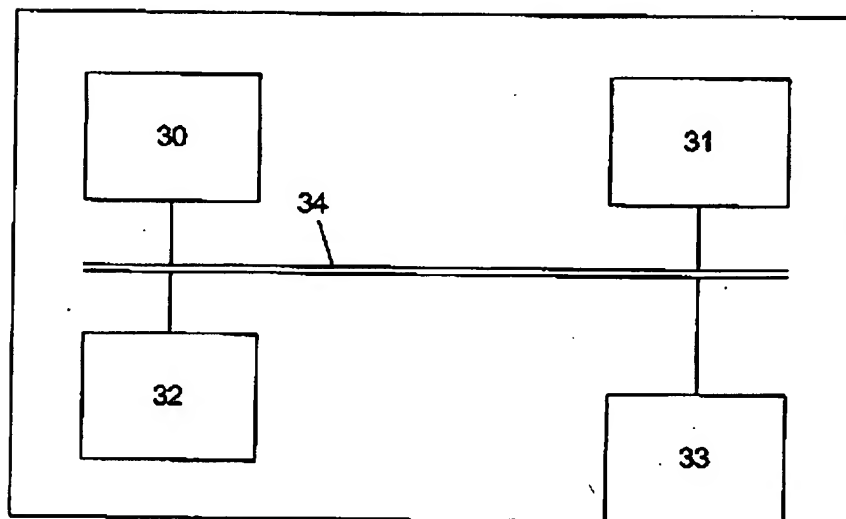
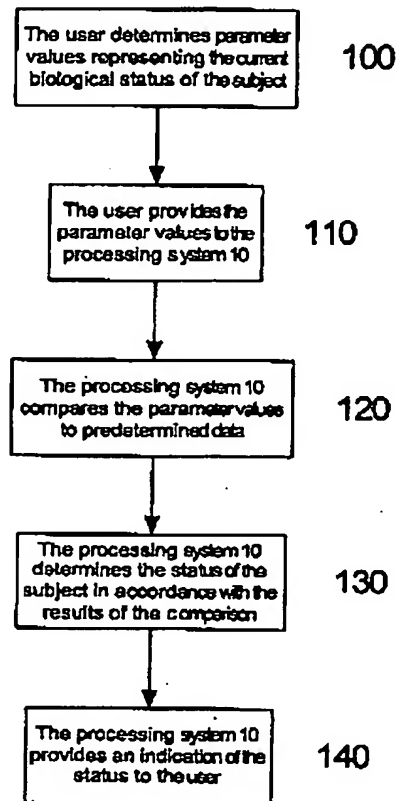


Fig. 4

**Fig. 2**

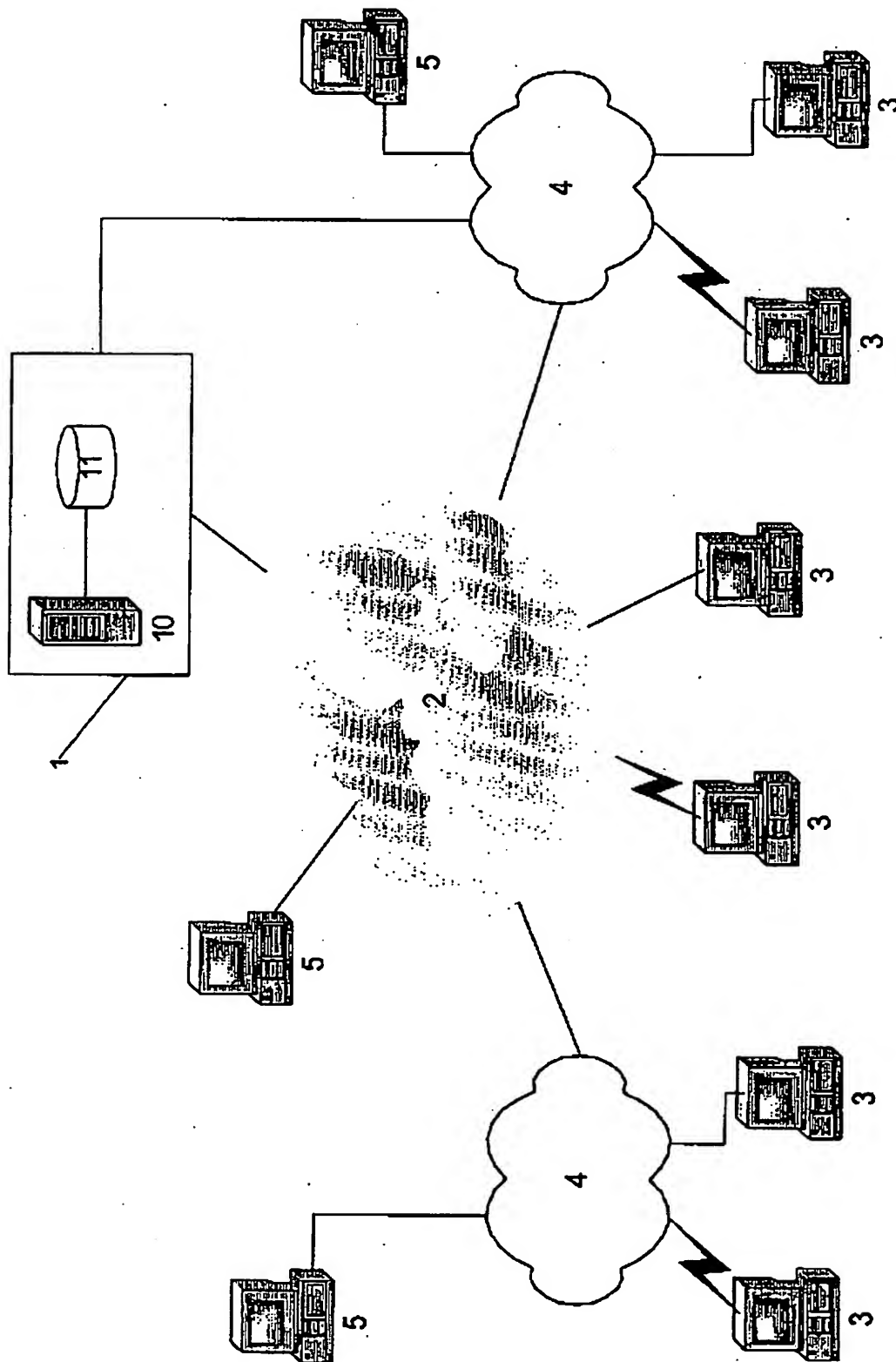
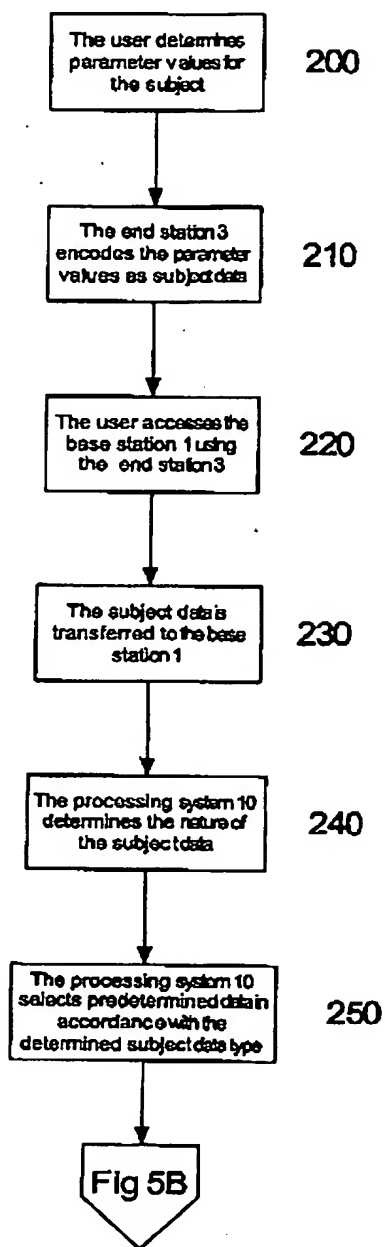
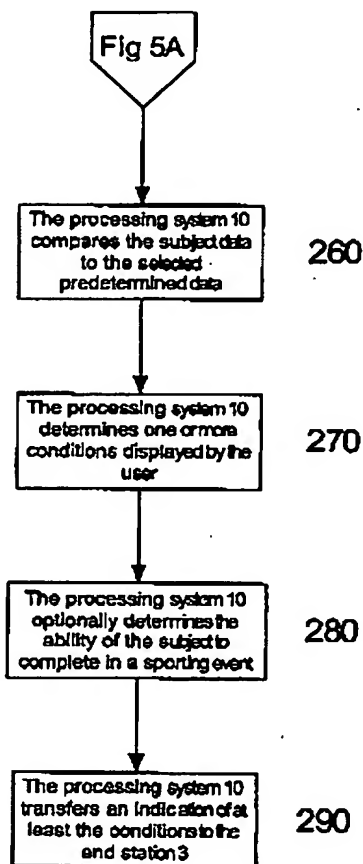
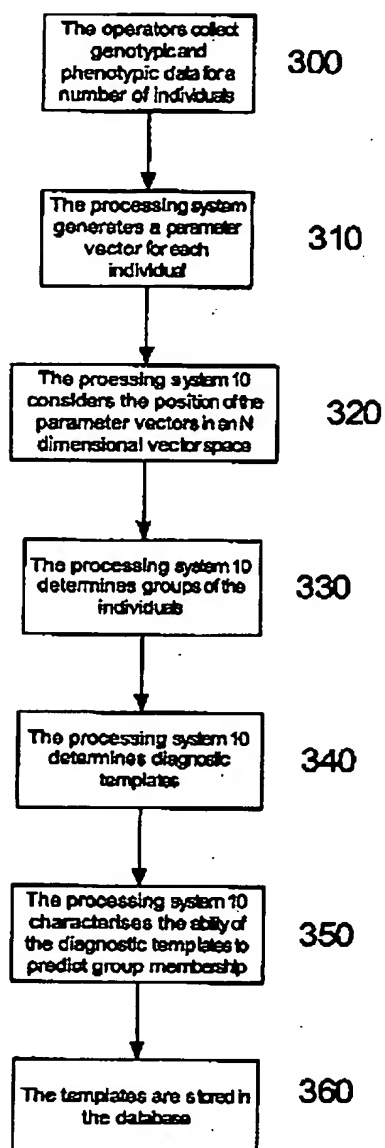
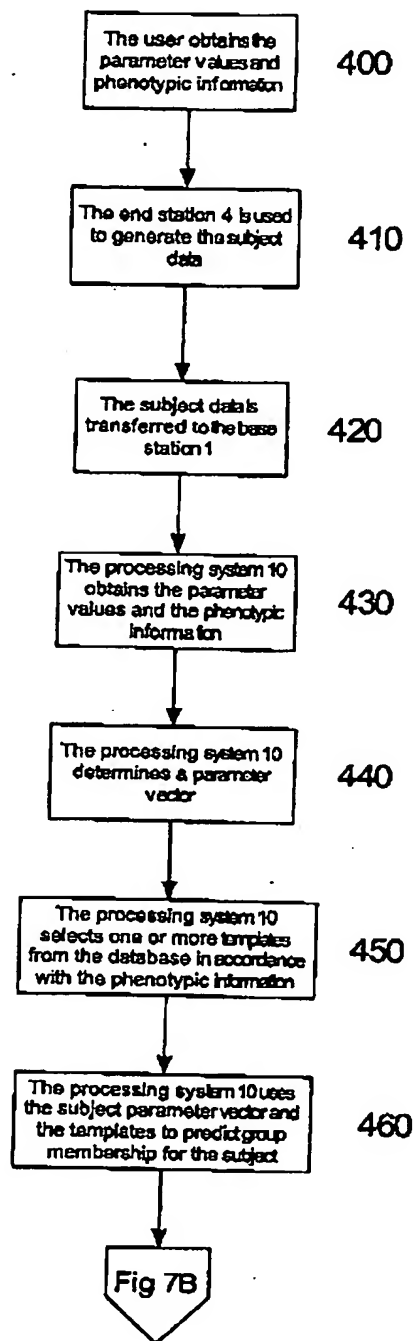


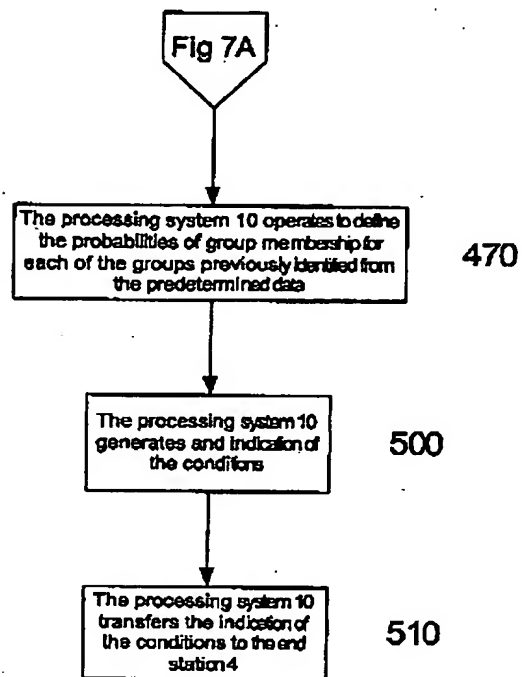
Fig. 3

**Fig. 5A**

**Fig. 5B**

**Fig. 6**

**Fig. 7A**

**Fig. 7B**

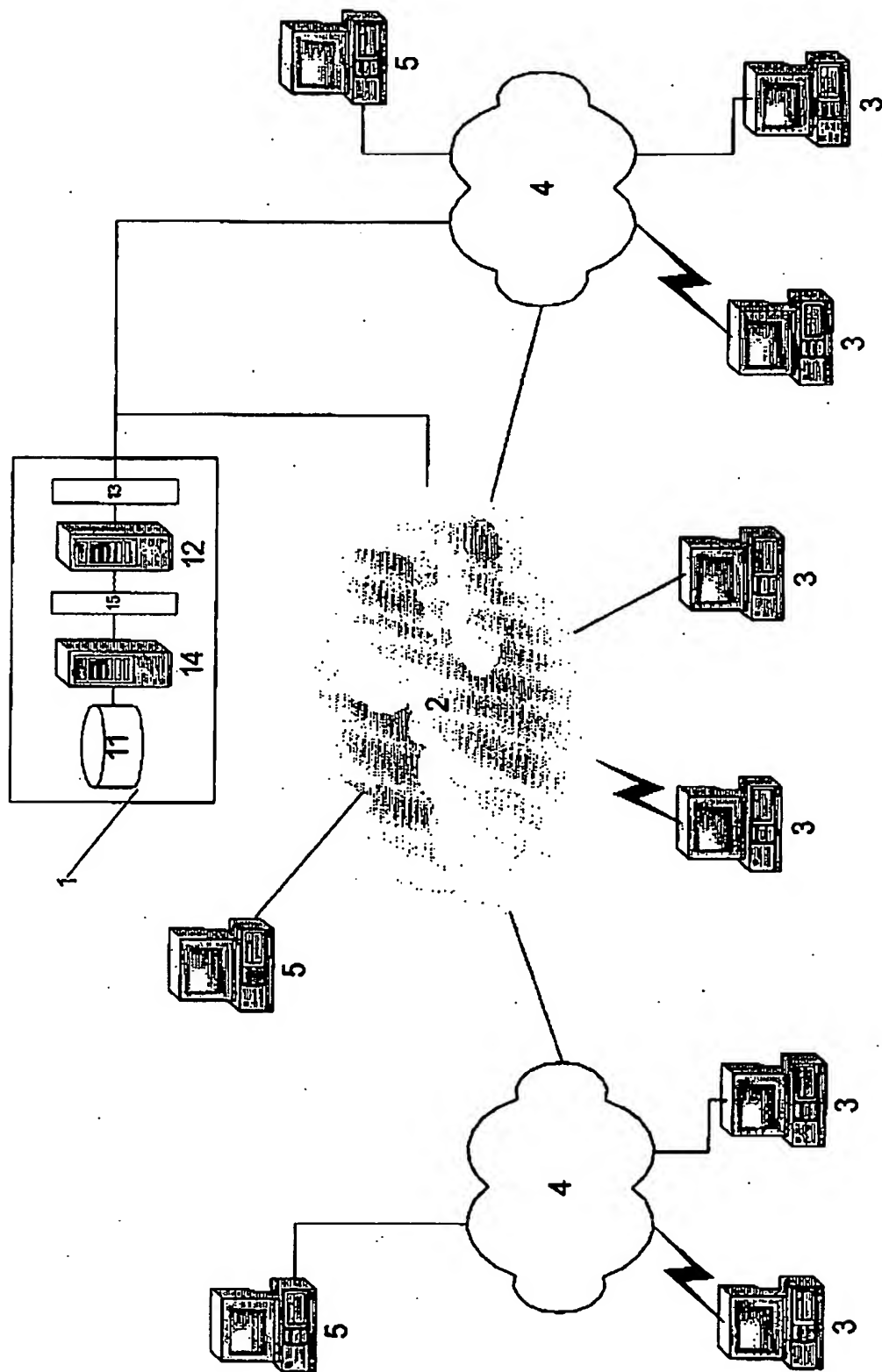
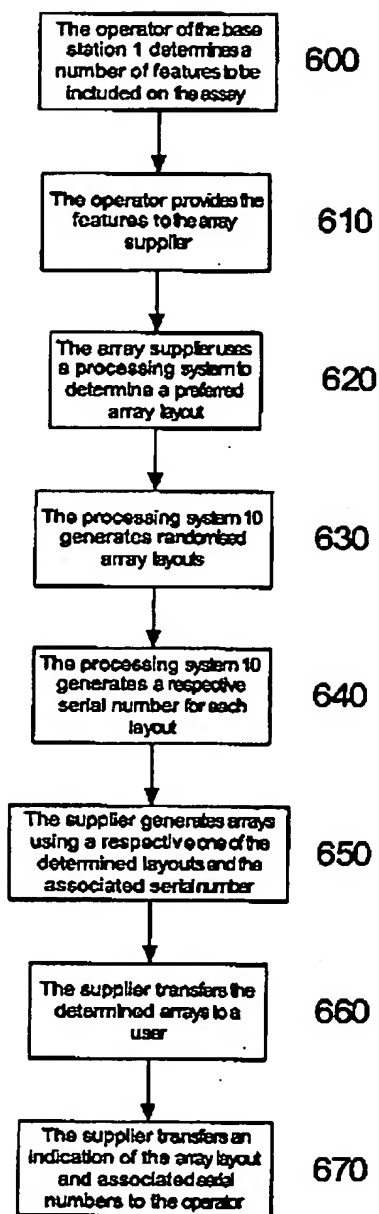
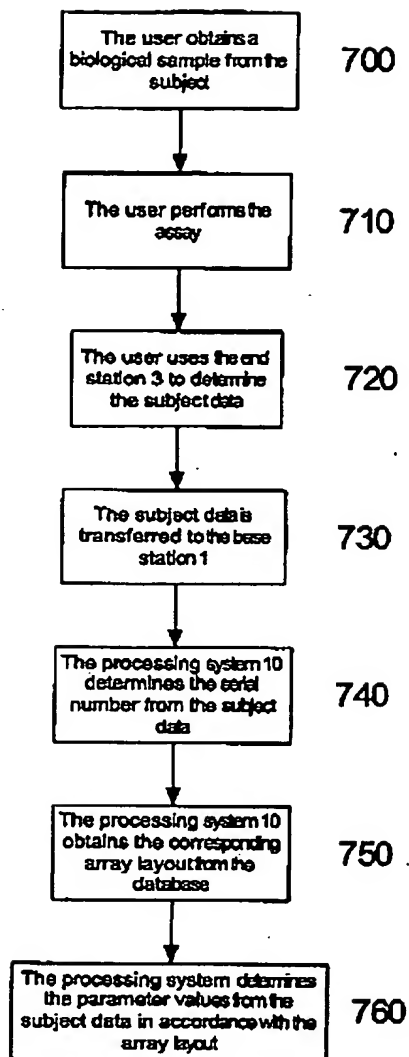


Fig. 8

**Fig. 9**

**Fig. 10**

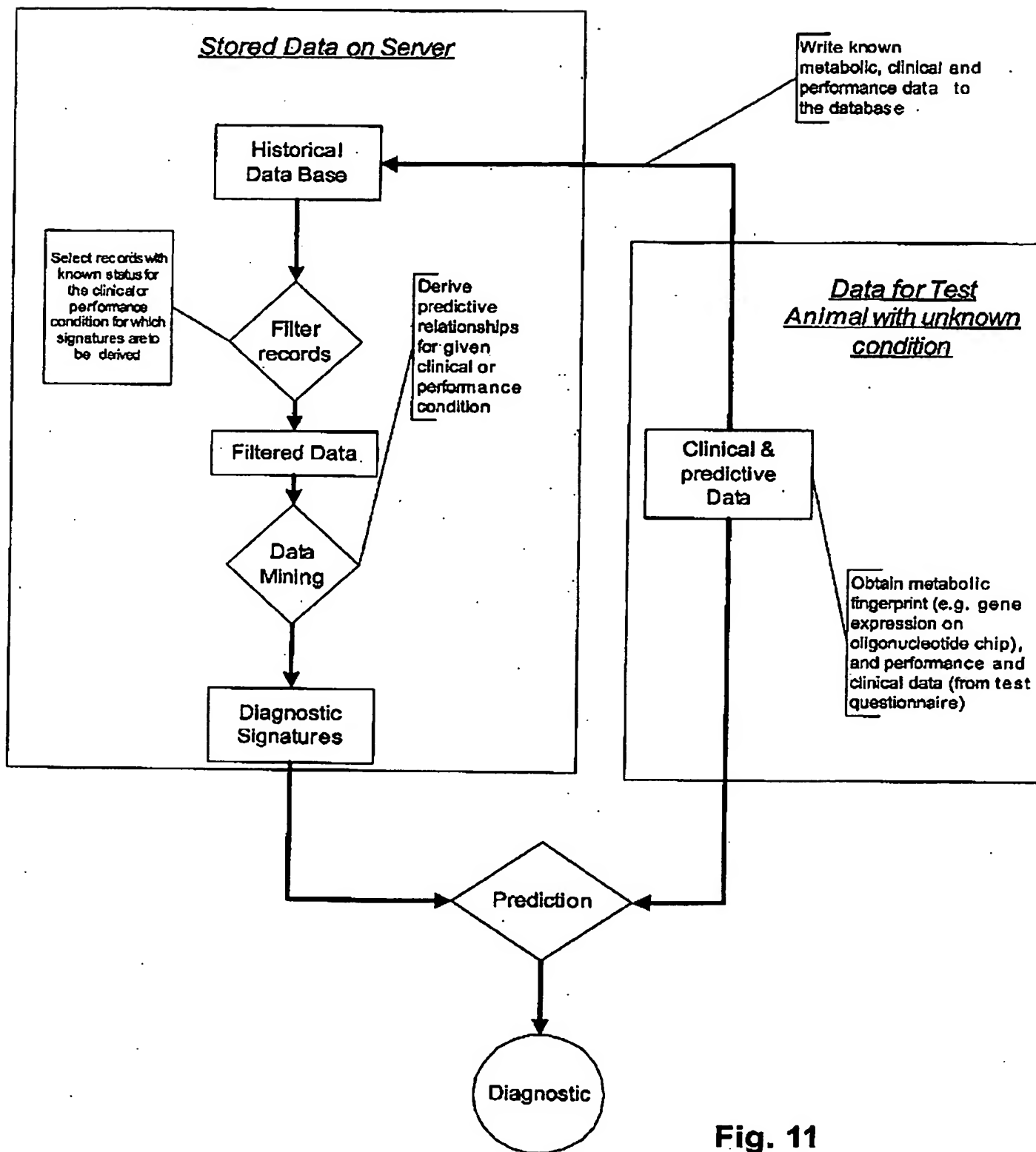


Fig. 11

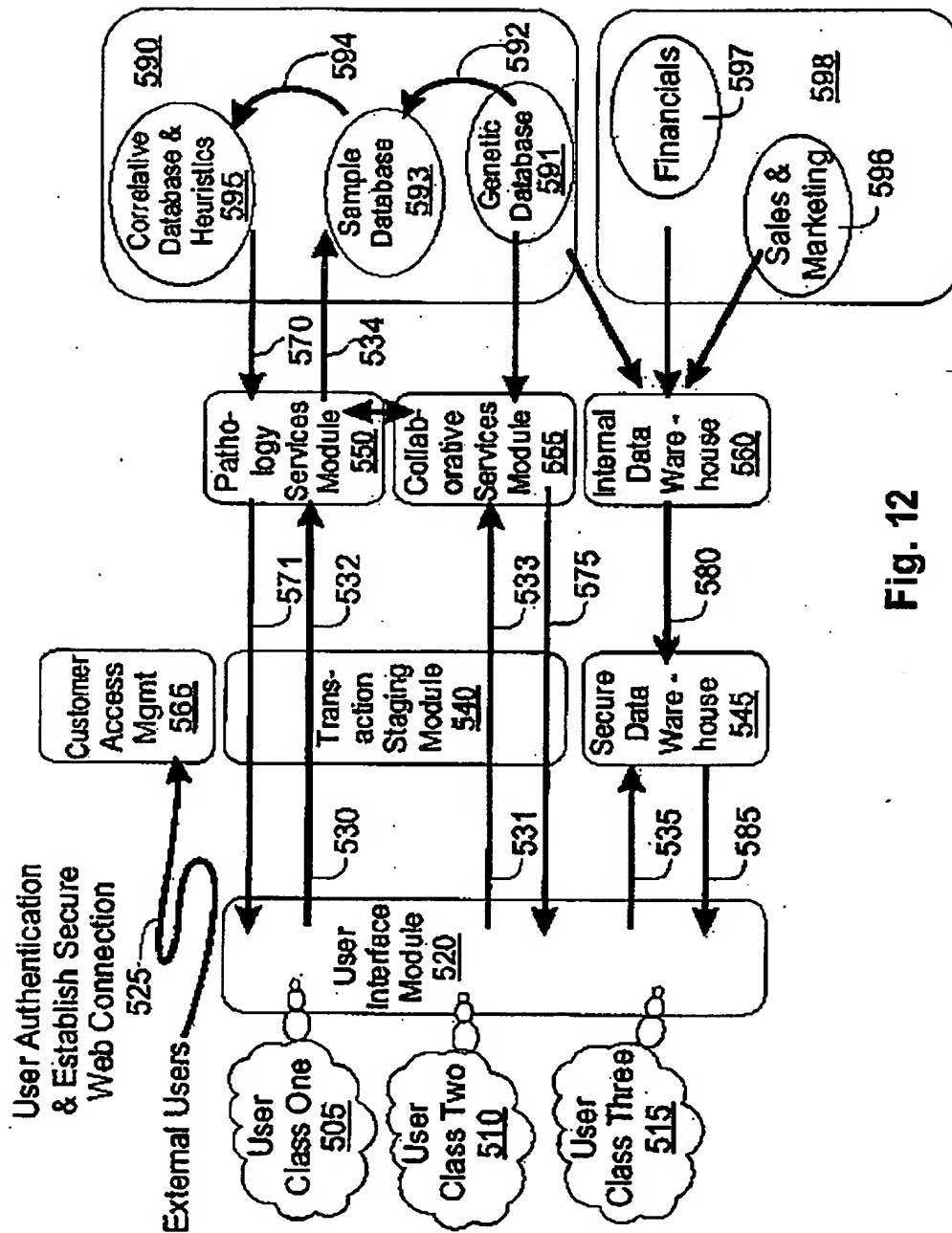


Fig. 12

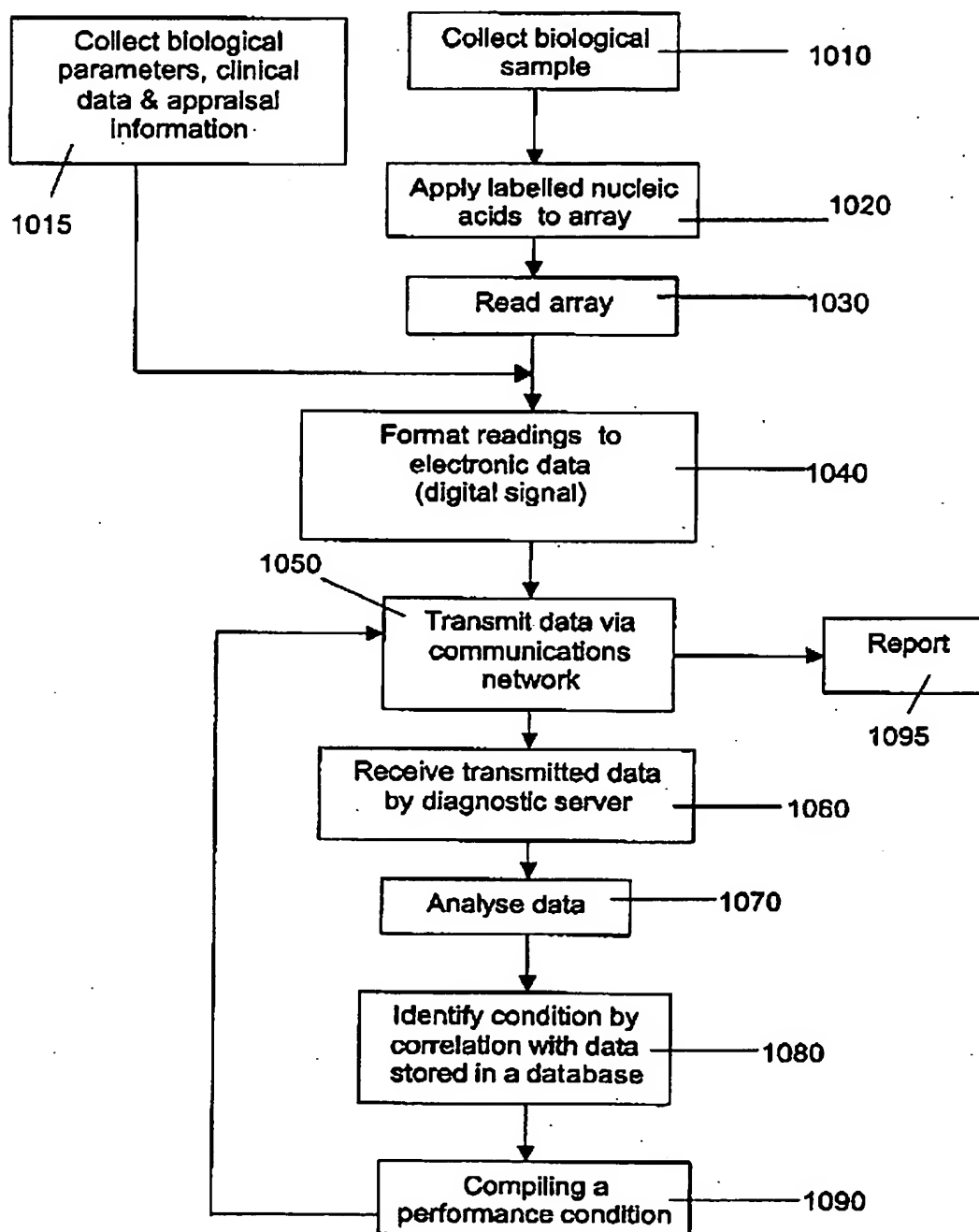


FIG. 13

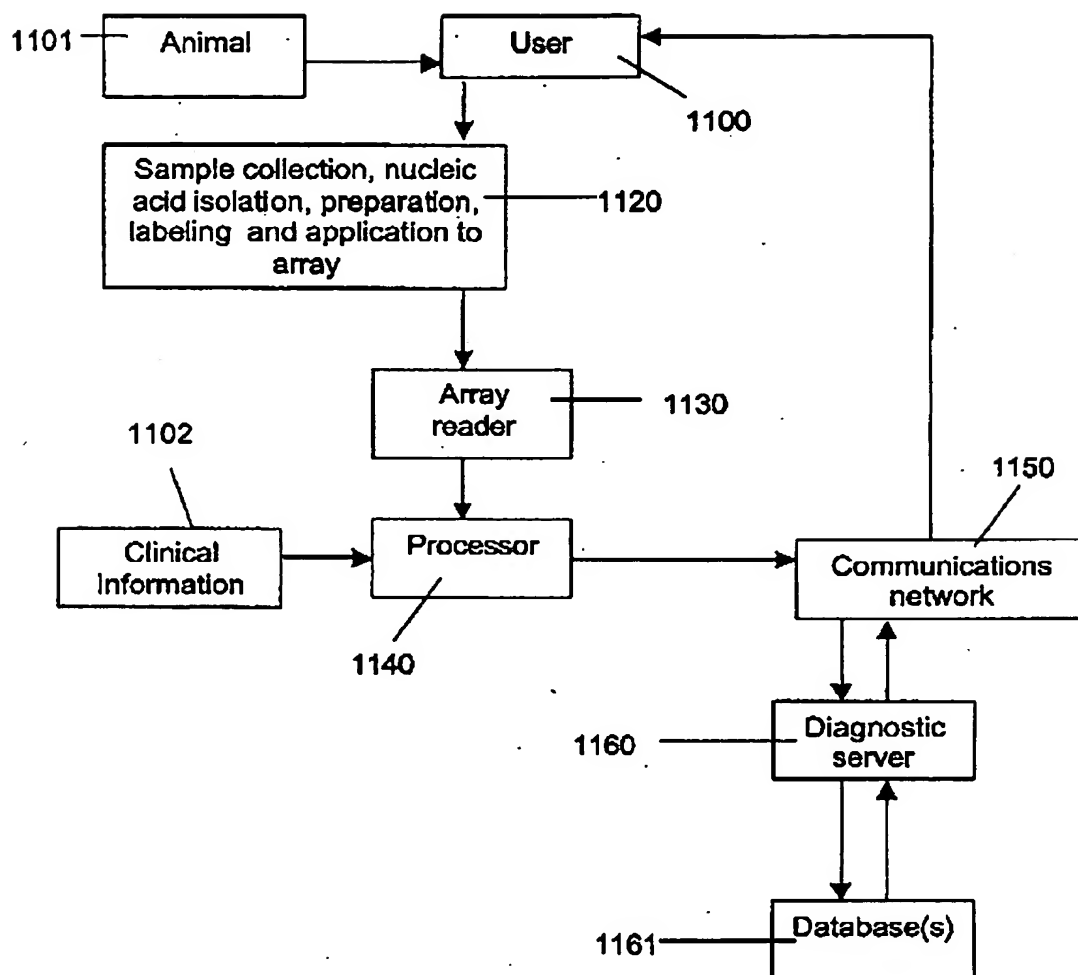


FIG. 14

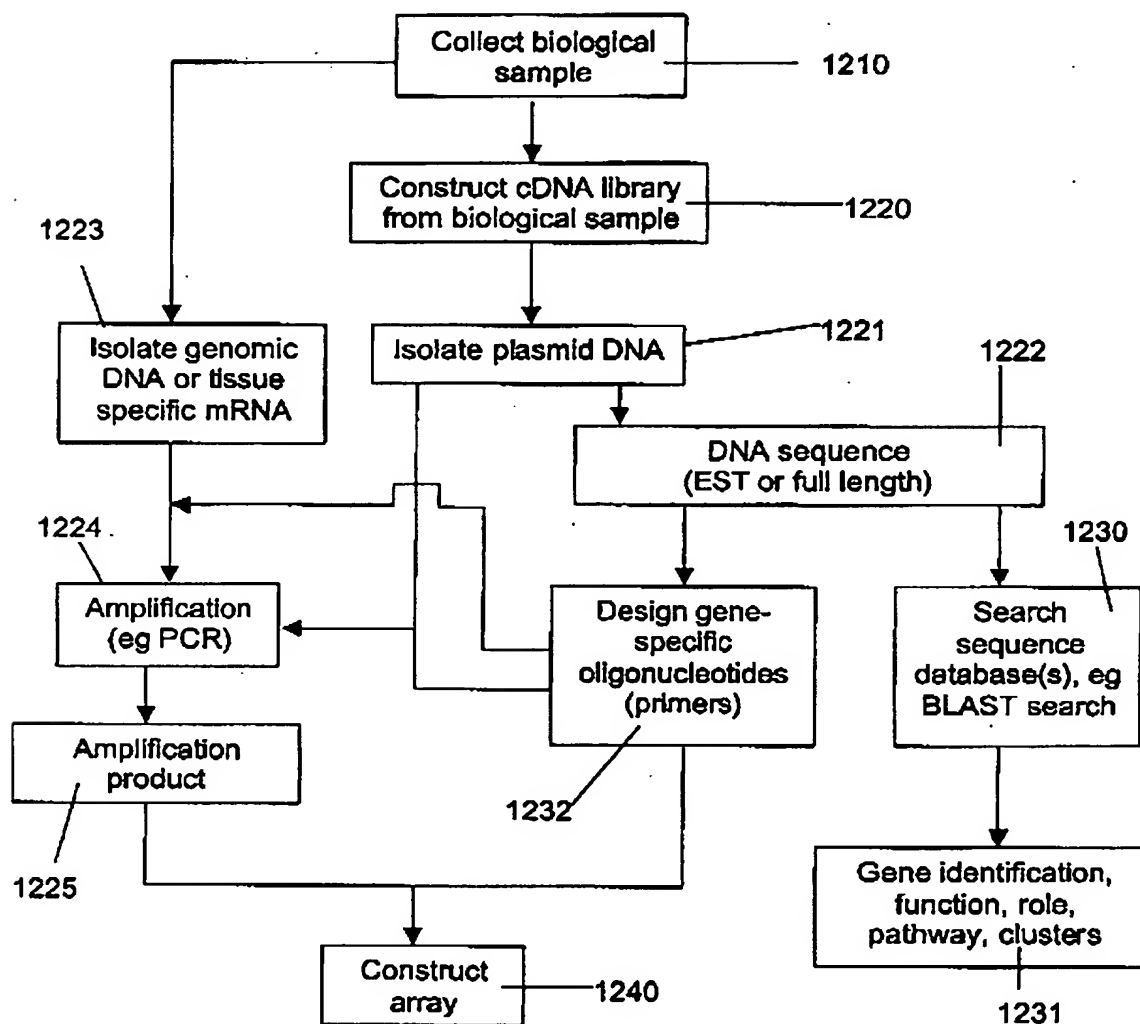


FIG. 15

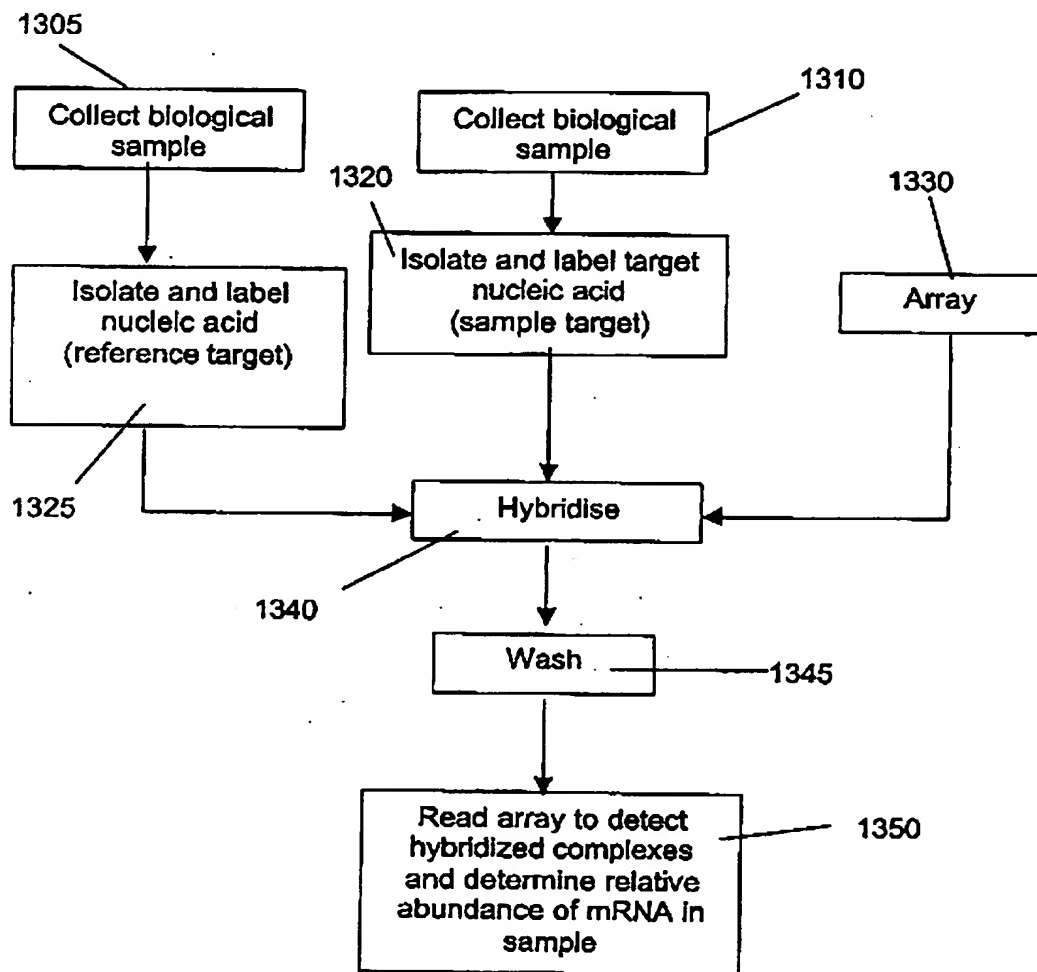


FIG. 16

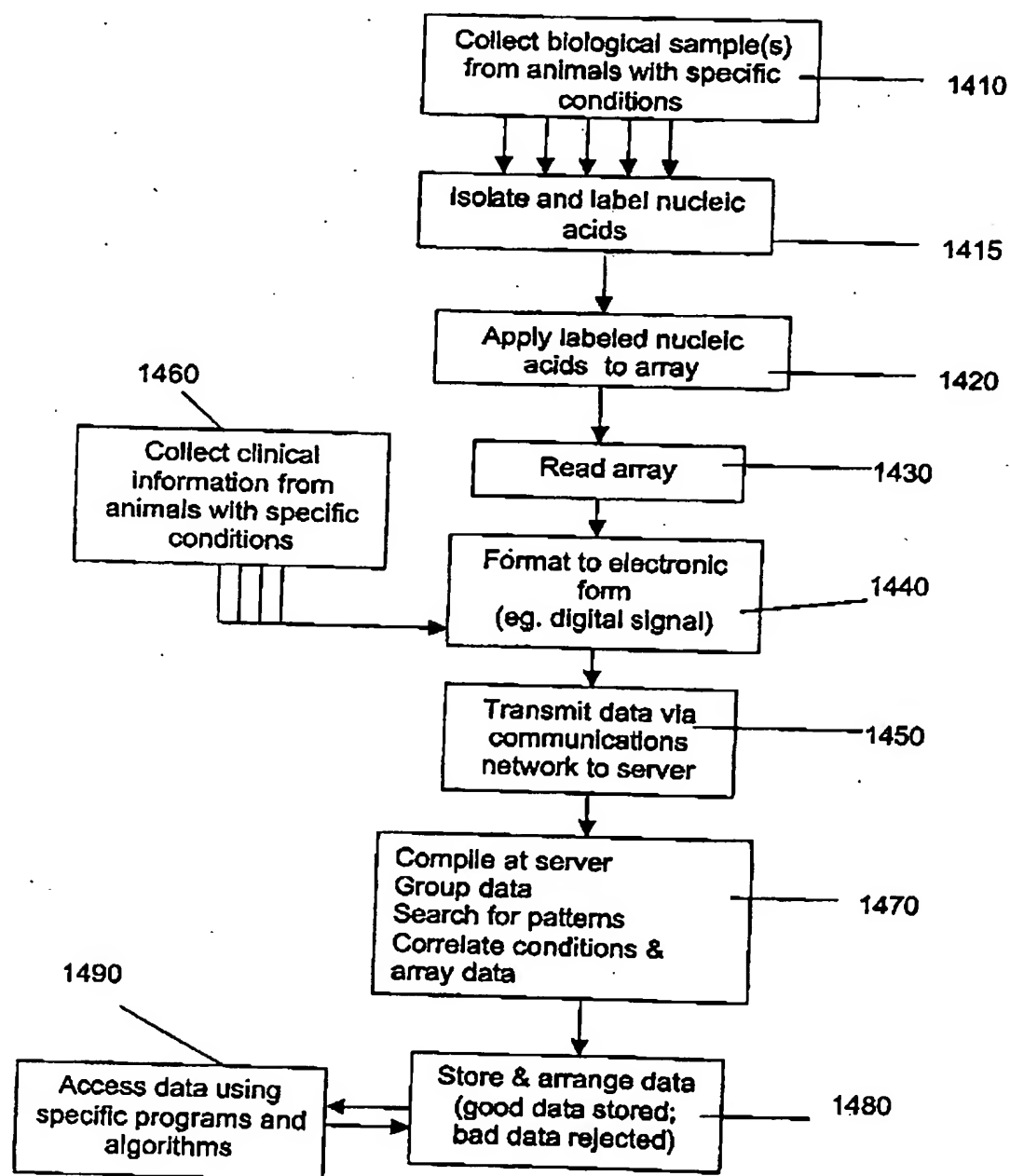


Fig. 17